

# UNCLASSIFIED

AD NUMBER
ADB224296
NEW LIMITATION CHANGE
TO Approved for public release, distribution unlimited
FROM Distribution authorized to DoD only; Specific Authority; 29 May 97. Other requests shall be referred to Commander, U.S. Army Medical Research and Material Command, Attn: MCMR-RMI-S, Fort Detrick, Frederick MD 21702-5012.
AUTHORITY
USAMRMC ltr, 4 Dec 2002

THIS PAGE IS UNCLASSIFIED

AD \_\_\_\_\_

CONTRACT NUMBER: DAMD17-95-C-5049

TITLE: A Transducer/Equipment System for Capturing Speech  
and Telemedicine Information for Subsequent  
Processing by Computer Systems

PRINCIPAL INVESTIGATOR: Benjamin Tirabassi

CONTRACTING ORGANIZATION: Technical Evaluation Research, Inc.  
Little Silver, NJ 07739-1162

REPORT DATE: April 1997

TYPE OF REPORT: Final

PREPARED FOR: Commander  
U.S. Army Medical Research and Materiel Command  
Fort Detrick, Frederick, MD 21702-5012

5-29-97  
DISTRIBUTION STATEMENT: Distribution authorized to DOD  
Components only (Specific Authority). Other requests for this  
document shall be referred to Commander, U.S. Army Medical  
Research and Materiel Command, ATTN: MCMR-RMI-S, Fort Detrick,  
Frederick, MD 21702-5012.

The views, opinions and/or findings contained in this report are  
those of the author(s) and should not be construed as an official  
Department of the Army position, policy or decision unless so  
designated by other documentation.

DTIC QUALITY INSPECTED 1

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE April 1997	3. REPORT TYPE AND DATES COVERED Final (15 Mar 95 - 14 Mar 97)		
4. TITLE AND SUBTITLE A Transducer/Equipment System for Capturing Speech and Telemedicine Information for Subsequent Processing by Computer Systems		5. FUNDING NUMBERS DAMD17-95-C-5049		
6. AUTHOR(S)  Benjamin Tirabassi				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  Technical Evaluation Research, Inc. Little Silver, NJ 07739-1162		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Commander U.S. Army Medical Research and Materiel Command Fort Detrick, MD 21702-5012		10. SPONSORING/MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES		19970527 120		
12a. DISTRIBUTION / AVAILABILITY STATEMENT Distribution authorized DOD Components only (Specific Authority). Other requests for this document shall be referred to Commander, U.S. Army Medical Research and Materiel Command, ATTN: MCMR-RMI-S, Fort Detrick, Frederick, MD 21702-5012.		12b. DISTRIBUTION CODE 5-29-97		
13. ABSTRACT (Maximum 200)  <p><b>Benchmarked speech capture data formed a corpus for the research to determine how the noise characterization information can best be used to improve speech recognition performance under tactical high noise conditions. The various speech capture improvement remedies investigated focused on a combination of methods that are matched to the FFT derived characterization of the tactical noise or medical sensor signals. These methods include stochastic canceling techniques for random (non-stationary) noise and algorithmic Feature Set subtraction for stationary noise. An important part of this investigation used tactical platform acoustic sound recordings under controlled conditions. The results of this experimentation was incorporated in speech processing algorithms for the remedy of stationary and non-stationary noise using post signal processing of Noise Feature Set Array Extraction techniques. Successful speech capture was demonstrated in noise environments up to 105 dBA using a combination of non-stationary and stationary noise post processing Noise Feature Set Extraction.</b></p>				
14. SUBJECT TERMS Speech Capture, Acoustic Benchmark, Stationary Noise Reduction, Speech Recognition in High Noise Environments, Spectral Feature Characterization, Telemedicine Channel Noise			15. NUMBER OF PAGES 63	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Limited	

## FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

\_\_\_\_ Where copyrighted material is quoted, permission has been obtained to use such material.

\_\_\_\_ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

X Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.


\_\_\_\_ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

\_\_\_\_ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

\_\_\_\_ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

\_\_\_\_ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

\_\_\_\_ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

 31 Mar 97  
PI - Signature Date

# A TRANSDUCER/EQUIPMENT SYSTEM FOR CAPTURING SPEECH AND TELEMEDICINE INFORMATION FOR SUBSEQUENT PROCESSING BY COMPUTER SYSTEMS

## Table of Contents

TITLE	PAGE
1.0 INTRODUCTION.....	1
2.0 RESEARCH SUMMARY FINDINGS.....	2
3.0 TECHNICAL OVERVIEW.....	4
<i>Software Programming.</i> .....	5
<i>API Routines.</i> .....	5
<i>Application Software.</i> .....	5
<i>Speech Capture System Architecture.</i> .....	6
4.0 TECHNICAL DISCUSSION.....	10
4.1 Noise Feature Set Subtraction Model and Algorithm Development.....	11
<i>SCAD Application Software and Program Interface.</i> .....	11
<i>Algorithm Refinement and Performance Testing.</i> .....	16
4.2 SCAD Hardware Architecture.....	16
4.3 Research Trials and Experiments.....	17
<i>90 dBA Synthesized Noise Trials.</i> .....	18
<i>100 dBA Synthesized Noise Trials.</i> .....	20
<i>100 dBA Field Noise Simulation Trials.</i> .....	21
<i>105 dBA Synthesized Noise Trials.</i> .....	22
5.0 CONCLUSIONS.....	24
6.0 RESEARCH FUTURE APPLICATIONS.....	26
 TABLES	 PAGE
TABLE 1: SCAD APPLICATION DESIGN.....	13
TABLE 2: SCAD TEST VOCABULARY.....	18

**A TRANSDUCER/EQUIPMENT SYSTEM FOR CAPTURING  
SPEECH AND TELEMEDICINE INFORMATION FOR SUBSEQUENT  
PROCESSING BY COMPUTER SYSTEMS**

**Table of Contents (continued)**

<b>APPENDED FIGURES</b>	<b>PAGE</b>
<b>FIGURE 1: SCAD APPLICATION FLOW DIAGRAM. ....</b>	<b>29</b>
<b>FIGURE 2: CONFIDENCE SCORES WITH 90 dB NOISE @ 500 Hz. ....</b>	<b>30</b>
<b>FIGURE 3: CONFIDENCE DELTA WITH 90 dB NOISE @ 500 Hz. ....</b>	<b>31</b>
<b>FIGURE 4: CONFIDENCE SCORES WITH 90 dB NOISE @ 1 KHz. ....</b>	<b>32</b>
<b>FIGURE 5: CONFIDENCE DELTA WITH 90 dB NOISE @ 1 KHz. ....</b>	<b>33</b>
<b>FIGURE 6: CONFIDENCE SCORES WITH 90 dB NOISE @ 2 KHz. ....</b>	<b>34</b>
<b>FIGURE 7: CONFIDENCE DELTA WITH 90 dB NOISE @ 2 KHz. ....</b>	<b>35</b>
<b>FIGURE 8: CONFIDENCE SCORE AND DELTA IMPROVEMENT FOR. ....</b> <b>VARIOUS BACKGROUND NOISE FREQUENCIES</b>	<b>36</b>
<b>FIGURE 9: CONFIDENCE SCORE AND DELTA IMPROVEMENT FOR. ....</b> <b>VARIOUS SIGNAL-TO-NOISE RATIOS</b>	<b>37</b>
<b>FIGURE 10: CONFIDENCE SCORES WITH 100 dB NOISE @ 500 Hz. ....</b>	<b>38</b>
<b>FIGURE 11: CONFIDENCE DELTA WITH 100 dB NOISE @ 500 Hz. ....</b>	<b>39</b>
<b>FIGURE 12: CONFIDENCE SCORES WITH 100 dB NOISE @ 1 KHz. ....</b>	<b>40</b>
<b>FIGURE 13: CONFIDENCE DELTA WITH 100 dB NOISE @ 1 KHz. ....</b>	<b>41</b>
<b>FIGURE 14: CONFIDENCE SCORES WITH 100 dB NOISE @ 2 KHz. ....</b>	<b>42</b>
<b>FIGURE 15: CONFIDENCE DELTA WITH 100 dB NOISE @ 2 KHz. ....</b>	<b>43</b>
<b>FIGURE 16: CONFIDENCE SCORES WITH 100 dB M1A1 BACKGROUND. ....</b> <b>NOISE</b>	<b>44</b>

**A TRANSDUCER/EQUIPMENT SYSTEM FOR CAPTURING  
SPEECH AND TELEMEDICINE INFORMATION FOR SUBSEQUENT  
PROCESSING BY COMPUTER SYSTEMS**

**Table of Contents (concluded)**

<b>APPENDED FIGURES</b>	<b>PAGE</b>
<b>FIGURE 17: CONFIDENCE DELTA WITH 100 dB M1A1 BACKGROUND. . . . .</b>	<b>45</b>
<b>NOISE</b>	
<b>FIGURE 18: CONFIDENCE SCORES FOR VARIOUS PRE-RECORDED. . . . .</b>	<b>46</b>
<b>FIELD NOISE TRIALS</b>	
<b>FIGURE 19: CONFIDENCE DELTA FOR VARIOUS PRE-RECORDED. . . . .</b>	<b>47</b>
<b>FIELD NOISE TRIALS</b>	
<b>FIGURE 20: CONFIDENCE SCORES FOR VARIOUS PRE-RECORDED. . . . .</b>	<b>48</b>
<b>FIELD NOISE TRIALS</b>	
<b>FIGURE 21: CONFIDENCE DELTA FOR VARIOUS PRE-RECORDED. . . . .</b>	<b>49</b>
<b>FIELD NOISE TRIALS</b>	
<b>FIGURE 22: CONFIDENCE SCORES WITH 105 dB NOISE @ 500 Hz. . . . .</b>	<b>50</b>
<b>FIGURE 23: CONFIDENCE DELTA WITH 105 dB NOISE @ 500 Hz. . . . .</b>	<b>51</b>
<b>FIGURE 24: CONFIDENCE SCORES WITH 105 dB NOISE @ 1 KHz. . . . .</b>	<b>52</b>
<b>FIGURE 25: CONFIDENCE DELTA WITH 105 dB NOISE @ 1 KHz. . . . .</b>	<b>53</b>
<b>FIGURE 26: CONFIDENCE SCORES WITH 105 dB NOISE @ 2 KHz. . . . .</b>	<b>54</b>
<b>FIGURE 27: CONFIDENCE DELTA WITH 105 dB NOISE @ 2 KHz. . . . .</b>	<b>55</b>
<b>FIGURE 28: SCAD CONFIDENCE SCORE COMPARISON WITH. . . . .</b>	<b>56</b>
<b>STATIONARY NOISE FEATURE EXTRACTION</b>	
<b>FIGURE 29: SCAD CONFIDENCE DELTA COMPARISON WITH. . . . .</b>	<b>57</b>
<b>STATIONARY NOISE FEATURE EXTRACTION</b>	

## **A TRANSDUCER/EQUIPMENT SYSTEM FOR CAPTURING SPEECH AND TELEMEDICINE INFORMATION FOR SUBSEQUENT PROCESSING BY COMPUTER SYSTEMS**

### **1.0 INTRODUCTION**

Research efforts successfully focused on improving automated computer speech capture and establishing a methodology for benchmark performance testing in various noise environments. Synthesized noise input frequencies and tape recorded tactical sound experiments were used to demonstrate the ability of the Fast Fourier Transform (FFT) real-time algorithm, as implemented on the Digital Signal Processor (DSP), to characterize and benchmark performance in various test noise environments.

Benchmarked speech capture data formed a corpus for the research to determine how the noise characterization information can best be used to improve speech recognition performance under tactical high noise conditions. The various speech capture improvement remedies investigated focused on a combination of methods that are matched to the FFT derived characterization of the tactical noise or medical sensor signals. These methods include stochastic canceling techniques for random (non-stationary) noise and algorithmic Feature Set subtraction for stationary noise. An important part of this investigation used tactical platform acoustic sound recordings under controlled conditions. The results of this experimentation was incorporated in speech processing algorithms for the remedy of stationary and non-stationary noise using post signal processing of Noise Feature Set Array Extraction techniques.

Two PCMCIA cards were fabricated that contain the FFT algorithms necessary to characterize the tactical noise or medical transducer signals. These portable and miniaturized devices can accomplish the sampling and identification of both stationary and non-stationary signals using a handheld computing device. Testing of the prototypes was accomplished with simulated tactical noise recorded signals for characterization analysis. The results showed both stable and repeatable attributes for the Speech Capture and Assessment Device (SCAD).

Application of the benchmark noise characterization techniques and remedies researched were successfully tested in a simulated tactical environment to determine the ability to characterize stationary noise using post Feature Set subtractive methods and to improve speech and medical signal information capture. Recursive testing and algorithm refinement was accomplished through a series of laboratory controlled and simulated tactical noise tests in support of the final miniaturized prototype device development. Successful speech capture was demonstrated in noise environments up to 105 dBA using a combination of non-stationary and stationary noise post processing Noise Feature Set Extraction.

This developmental research culminated in the design and productization of a superior noise characterization and speech capture system embodied on a PCMCIA card. This SCAD



system embodies the successful results of this research in a miniaturized and refined product suitable for the soldier. The speech capture system represents the leading edge technology in voice interaction under very noisy battlefield conditions and simultaneously resulted in a cost-effective commercial product with mobile and vendor kiosk applications.

## 2.0 RESEARCH SUMMARY FINDINGS

Several important performance generalities to support the Feature Set Array Extraction theory are exhibited in the captured tactical sound data. Evidence of the fact that each particular tactical sound has a stationary spectral component is of paramount importance to this research effort. The unique identification of the sounds resides in the relative power intensity associated with the FFT computed information allocated to frequency "bins" across the 50-4000 Hz spectrum. These power level characteristics remain stable and repeatable for the sampled data and show a smoothing effect with increasing number of multi-frame averaged data.

Multi-frame capture averaging of the spectral characteristics exhibited an interesting phenomena which correlates with the postulate: that subtraction of the averaged (stationary) noise characteristics at the feature level is desirable. The captured sound data definitely shows that the subtraction of instant analog noise data (front end noise cancellation) would extract too much of the desirable signal power over the spectrum, and adversely affect speech recognition accuracy. However, subtraction of the averaged noise spectrum power that has been characterized as a Feature Set would be more deliberate in canceling the stationary components of the noise. Exactly how much multi-frame averaging of the FFT relative power would produce optimum results was part of the study effort.

The remedies postulated to improve the capture of speech (acoustic data) in a tactical noise environment are comprised of two complementary algorithm controlled processes. It is taken as a given that the noise (N) that is mixed with the signal (S) has two components:  $N_r$ , which is a random noise component; and  $N_s$ , which is the stationary noise component. First, the elimination of the non-stationary (random) noise is accomplished through stochastic methods applied to the signal plus noise (S+N) data after the digitization of the analog signal is complete. Second, the removal of the stationary noise component is accomplished through feature extraction using captured (S+N) and noise (N) Feature Set processing. It should be noted that both of these processes circumvent the "front-end" analog cancellation approach which is known to extract (filter) too much of the desired signal in high noise and low  $S/(S+N)$  ratio environments.

Implementing the non-stationary noise ( $N_r$ ) reduction is a straight forward process, implemented by applying stochastic averaging and variance mathematical algorithms to the digitized (S+N) captured audio. Eliminating the stationary noise ( $N_s$ ) proves to be more challenging since it relies upon the ability to properly characterize the noise. Once the noise is identified it is necessary to codify the characteristics in a form that is identical to the Feature Set used for automatic recognition processing. Early research in this program was successful in the characterization of the stationary noise ( $N_s$ ) using a real-time FFT capture DSP implementation.

Those (N<sub>e</sub>) characterization investigations exhibited the fundamental qualities of stability and repeatability which encouraged the next stage of algorithm development and testing.

The developed noise Feature Set subtraction method for improving speech capture was successfully demonstrated in 90 dBA, 100 dBA, and 105 dBA noise environments using the SCAD. It was determined that a set of consistent coefficients for feature subtraction algorithms proved optimal under all tested conditions of variable frequency, dB level and signal to noise ratio. Similar improvements in speech capture accuracy were recorded for individual stationary noise frequencies as well as complex harmonic rich noise environments containing tactical battlefield platforms. Stochastic cancellation of random noise continued to perform as expected in the presence of stationary noise feature subtraction algorithms. The developed SCAD simultaneously processes both noise characterization and speech recognition algorithms to improve the capture of speech in very high level noise environments containing both random and stationary noise.

Synthesized fixed frequency stationary noise at 100 dBA and 105 dBA was used to determine optimum algorithm coefficients to expand the previous testing conducted at 90 dBA. Baseline speech capture performance was compared with noise feature subtraction methods to show improvement achievable at various background noise levels. Speech input level was researched as an independent variable and was examined in terms of signal to noise ratio to judge its effect. Optimum speech capture resulted when the signal was approximately at the same audio level (or less) than the noise under most conditions, which reinforces the knowledge that very loud speech is distorted and undesirable.

These findings represent a breakthrough in speech processing, and especially in speech recognition, that previously required a 5 to 10 dB level of speech signal higher than the noise for speech intelligibility. The noise feature extraction developed algorithms, when applied to the baseline speech recognition capability, continued to perform well at equal levels of speech and noise audio levels. Particular words in the speech recognition vocabulary, that did not contain plosive sounds, did not fare as well at the 105 dBA noise level.

Improved speech capture was demonstrated by the synthesized stationary frequency testing at 500 Hz, 1 KHz and 2 KHz. More improvement was noted at the higher frequencies due to the physical and electrical filter break points that are inherent in the speech capture hardware and channel components below 1 KHz. Testing above 2 KHz was not conducted since greater improvement in performance was predicted by the empirical data. Successful testing in the presence of M1A1 Tank field recorded complex spectral noise with range 50 Hz to 4 KHz at

100 dBA proved this assumption correct. Speech capture improvement in the presence of complex M1A1 noise was demonstrated during testing to be similar to the average response data observed for the synthesized range of frequencies. Speech capture accuracy and word recognition improved 10% in the 100 dBA M1A1 noise environment and 15% on average in the presence of M113, Jet Aircraft and Helicopter noise.

It is noted that the improvement in speech capture accuracy derived from stationary noise feature extraction techniques is in addition to the already defined baseline. The baseline already has significant performance improvements inherent in the noise canceling microphone and the statistical random noise elimination algorithms. The baseline configuration permits the speech recognizer to continue operation in noise environments above 90 dBA to 105 dBA. The addition of the stationary noise feature extraction techniques extends operations up to 105 dBA and also improves accuracy by 15% or more above 90 dBA, dependent upon the spectral content of the noise.

### 3.0 TECHNICAL OVERVIEW

The design and packaging approach successfully achieved the miniaturization goal for the SCAD in a PCMCIA standard card product. This miniaturized product can be used as a "front end" to standard Army fielded products (e.g., computers, intercoms, radios, etc.) providing hands-free voice (or medical sensor) interaction under tactical conditions. The miniaturization also reduces power requirements to a fraction of a watt so that it may be powered directly from the host and require no separate power source. This will lighten the soldier's load and provide a commercially viable product to "front-end" small even palm-top computers with their limited battery resources.

The refined speech capture system contains the speech processing adaptability algorithms that result from task research which dynamically changes the speech capture parameters and adapts to the acoustic and signal channel environment. These algorithms are based upon the research conducted for benchmark noise characterization and the successful remedies developed as a result of laboratory and field experiments. Benchmark performance assessment is automatically calculated and speech capture algorithms parameters modified in real time to adapt to the "sensed" acoustic and environmental conditions. The speech capture system incorporates both random and stationary noise characterization algorithms to assess the present set of acoustic conditions. Additionally, the innovative noise reduction adaptive signal processing and speech parameterization techniques are incorporated for more robust speech recognition performance given the instant noise conditions. The algorithms incorporated in the SCAD demonstrate immunity to a broad range of acoustic and transmission media noise conditions which include concentrations of characterized tracked vehicle and battlefield noise.

The development of the SCAD has been divided into two efforts: the design phase and the fabrication phase. Design effort has been further subdivided into software programming and hardware architecture development subtasks.

## *Software Programming*

The software development effort for the prototype SCAD involved two sequential steps. First, hardware compliant Application Program Interface (API) modules were built so that individual SCAD functions (e.g., Get Speech Utterance) could be called as library routines rather than using redundant DSP coding. The second step was to build an application program that would capture signal plus noise utterances, convert them into Feature Set Arrays, then subtract any intruding noise in the signal by comparing the Feature Set Array to the noise only Feature Set Arrays that are dynamically captured. All API and application software development is written in American National Standards Institute (ANSI) compatible "C" language.

### *API Routines*

The following is a list of developed API routines. The routines are available as a library of object modules that are linked with the application.

<u>API</u>	<u>Description</u>
Open	Initializes the SCAD hardware
OpenVoc	Opens a user defined Feature Set Array file on the host hardware
Record	Makes a digital audio recording of an utterance using the SCAD endpoint detection algorithm and stores it as a Feature Set Array of elements
Play	Plays a digital recording
Get Speech Utterance	Obtains a recognition response from the SCAD
Close Voc	Closes the Feature Set Array file that is currently open
Close	Shutdown the SCAD

### *Application Software*

Application software was developed to demonstrate the advanced technology being inserted into the SCAD. Software development began by first incorporating the developed API's into callable routines that are recognizable by the SCAD hardware. Each routine performs a specific task and returns the appropriate confirmation or error codes to inform the user regarding the functional state of the SCAD unit. Executable code has been successfully compiled to demonstrate each of the API routines.

Of paramount importance to the performance validation of the SCAD HW/SW combination is ultimately how well it can recognize one utterance out of a list of words in the presence of varying noise environments. To assess a tangible variable to each speech utterance trial, the Get Speech Utterance API has been architected to return the top 4 most likely candidates recognized out of the list of stored utterances on-board. Each of the 4 are listed and assigned a numeric value from 1 to 1999 where the lower score means a higher confidence was

assigned by the SCAD and that it was properly recognized. Likewise, a delta value is calculated as the difference between the 2 best response scores and assesses how well the number 1 choice was discernible from its next closest choice. The lowest score and largest delta would be considered the best choice for a valid utterance.

The goal of the software application is to capture an analog signal plus noise utterance in a given environment. Many of the advanced filtering techniques incorporated into the SCAD hardware will eliminate most of the random noise associated with the captured utterance resulting in a digitized Feature Set Array of signal plus stationary noise. The SCAD then analyzes a noise only Feature Set Array that occurs just prior to a signal plus noise utterance capture. Dynamically subtracting a carefully chosen power normalized noise only Feature Set Array in "real-time" mode from the captured signal plus noise Feature Set Array is then processed to determine score and delta.

The output of the application software effort is two-fold. First, a preprocessed match of the signal plus noise Feature Set Array will be compared with stored utterance templates. The results of the match will be in the form of a recognition score and delta value of the top 4 best choices. The SCAD then analyzes the Feature Set Array, previously mentioned, and "subtracts out" the noise from the signal plus noise utterance. The SCAD then performs a second Feature Set Array comparison of the adapted utterance against the stored templates. The goal output, assuming the stored vocabulary Feature Set Arrays were captured in a benign environment (no noise), would be a lower recognition score (higher confidence) and larger delta (more discernible) for the adapted Feature Set Array of signal plus noise when compared with the unadapted set.

### *Speech Capture System Architecture*

Prototype SCAD hardware has been architected as a low power multi-chip module laminate (MCM-L) incorporated in a Type II PCMCIA form factor. The miniaturized form factor provides a portable and PC compatible solution for capturing and analyzing audio samples. The PCMCIA card MCM has been built to include integrated chip A/D and D/A conversion for a full 16-bit accumulator (96 dB dynamic range equivalent) voiceband analog interface to process the audio input and output, enough Flash RAM to store captured speech and noise Feature Set Arrays, and an embedded Analog Devices 21ADSP55a Mixed DSP for real-time FFT computation of characterized speech and noise samples.

The prototype SCAD conforms to the IEEE PCMCIA Version 2.0 standards for PC hardware and software interface requirements. User interface microphone and speaker connections are provided through standard audio DIN connectors. On-board programmable signal level, impedance matching and bandpass filters are capable of interfacing with a wide range of commercial and military I/O devices, such as; dynamic microphones (typical output range 0.1 to 50.0 mV), noise-canceling microphones (typical output range 1.0 to 5.0 mV), headsets and speakers (4-16 ohms). Through the use of an external Push-to-Talk (PTT) switch, the user can

control the on-set of a speech utterance to differentiate it from the automatic noise sampling and characterization mode of operation.

Making use of a flexible mixed DSP and Field Programmable Gate Array (FPGA), the SCAD hardware will be able to mature as improvements to the speech processing algorithms evolve with field testing. On-chip Random Access Memory (RAM), which is external EPROM programmable, allows for easy code revision at low production costs without a commitment to one-time programmable Read Only Memory (ROM) devices. EPROM loading also supports dynamic loading and reconfiguration during operation which will allow for the real-time field modification of signal plus noise Feature Set Arrays in host configurations.

The research developed algorithms and methods showed the feasibility of using an automated benchmark technique to quantify performance of a speech capture system. Early research applied these methods to other noise environments in order to develop a corpus of data. It has been shown that an understanding of the type of characteristic noise derived from the automated FFT algorithm is fundamental to eventual speech capture improvement. Specifically, data was gathered for a variety of input transducers to assess their effect on speech intelligibility in noise environments. The benchmarking algorithms were then used to assess these and other previously proposed remedies. These alternative and combined remedies were subject to experiment and data was taken to quantify improvement of these remedies in the presence of various characteristic noise environments. An important aspect of these experiments explored the physical practicability and comfort issues for physically active soldiers.

Experiments made use of the performance benchmarking techniques and noise characterization tools developed. Use of these tools were essential to the testing of different speech capture appliances and intelligibility enhancement. These Articulation Index (AI) performance and FFT characterization tools benchmark the various remedies giving repeatable and reliable intelligibility data during the laboratory and simulated field trials. To support these trials required the incorporation of the automated FFT and performance benchmark algorithms into a portable and self-contained prototype device.

Technical Evaluation Research Inc (TERI) fabricated several prototype SCADs used during the experiments and trials. These SCADs gathered the benchmark data in highly mobile situations to accurately determine cause and effect in simulated tactical situations. Front-end design of the SCAD permitted rapid modification and exchange of key voice channel devices (like transducer microphones) during the experiment. SCADs incorporated the automated algorithm driven noise characterization capability, and in real-time modified the recognition parameters to optimally adjust to these environmental noise characteristics. The combination of hardware and software remedies provided a comprehensive testbed to further refine the physical suitability of the human spoken interface in the presence of high noise.

Experiments were conducted in both laboratory and synthesized tactical environments through coordination with Aberdeen Proving Grounds, MD. These trials made maximum use of the SCADs to document the performance and attributes of the various speech capture improvement remedies. These remedies are suitable to: an individual soldier running in MOPP uniform, a computer operator in a moving tracked vehicle, an operator in a moving wheeled vehicle, a helicopter crew member or a crew member firing a tank gun or howitzer. Simulated noise environments were used to generate the data corpus which are documented together with conclusions regarding the achieved speech capture improvement and recommended algorithms.

The culmination of the research is marked by the redefinition of a successful speech capture system for automated processing and further miniaturization. It is important to the commercialization of this product that the implementation be small and inexpensive. This SCAD weight, power, and cost factors are also very important to the individual soldier. A singular low power and lightweight design has been developed that can be universally applied in different tactical situations and commercial high noise environments.

The exploratory research approach sought to parameterize the speech signal, provide measurements, and provide superior speech capture while minimizing the effects of noise and interference. TERIs active participation in state-of-the-art voice processing represented a unique opportunity to extend these recognition algorithms, which have been demonstrated to be speaker and channel independent, to include noise immunity.

The independent speaker environment also contributes to the successful continuance of speech intelligibility under stress and other media imposed conditions because of the algorithm inherent channel and noise statistical correlation rejection qualities. Contributing to the robust nature of TERIs speech processing algorithms is the continued research of the adaptability feature which dynamically changes capture parameters to adapt to existing acoustic and channel environment. This adaptive feature is selective and permits dynamic algorithm tailoring to maintain speech accuracy under stress conditions and high levels of background acoustic noise. These algorithms have been developed and tailored for specific military use in a broad range of acoustic conditions to include remote communication, armored vehicles, and aircraft.

Of significance from this research is the development of a benchmark performance methodology and noise conditions that are documented for several anticipated environments using different acoustic noise and channel media conditions. This benchmark methodology and documented noise conditions represent a repeatable and stable baseline for the research effort to determine speech capture algorithm effectiveness in the anticipated environments. An important part of the research effort is dedicated to the methodology and a description of the techniques proposed for performance testing utilizing a sample of the anticipated noise environments. Several prototype SCAD systems have been designed and assembled to include algorithm enhancements, transducers, and necessary ancillary devices for speech intelligibility enhancement and demonstration in high noise environments. The algorithms demonstrate immunity to a broad

range of acoustic and transmission media conditions which include concentrations of acoustic noise in all anticipated regions of the spectrum which is significant in the human voice band.

It was determined how the noise characterization information can best be used to improve speech recognition performance under tactical conditions. The various remedies investigated focused on a combination of methods that are matched to the FFT derived characterization of the tactical noise or medical sensor signals. These methods include stochastic canceling techniques for random (non-stationary) noise and algorithmic feature subtraction for stationary noise. An important part of this investigation used tactical platform acoustic sound recordings under controlled conditions. The results of this experimentation is the demonstration of superior speech capture in the presence of stationary and non-stationary noise using post signal processing Feature Set Array subtraction techniques.

Two prototype PCMCIA cards that contain the FFT algorithms were used to characterize the tactical noise or medical transducer signals. These portable and miniaturized devices can accomplish the sampling and identification of both stationary and non-stationary signals using a handheld computing device. Testing of the prototypes was accomplished with tactical noise signals for characterization analysis. The results showed both stable and repeatable attributes.

Comprehensive laboratory testing was conducted to evaluate the prototype algorithms and transducer/equipment performance using high noise synthetic and field captured noise environments. Application of the benchmark techniques and remedies researched are tested in a simulated tactical environment to determine the ability to characterize stationary noise and improve speech and medical signal information capture. SCAD testing of in-laboratory performance trials was used to collect data for the optimization of the algorithms for the final miniaturized prototype device development effort. Recorded field noise for the M1A1, M113, Helicopter, Jet, etc. was conducted to confirm predicted SCAD product performance. This course of action was considered to be more effective, to benchmark the SCAD performance under laboratory controlled conditions, using repeatable synthesized and field noise environment recordings.

Synthesized stationary noise at 100 dBA and 105 dBA is used to determine the optimum algorithm coefficients and threshold settings for the SCAD. Research testing for a range of variable settings indicated the optimal performance was obtained using approximately the same noise feature subtraction coefficient as determined at lower noise levels (e.g., 90 dBA). Improved speech, captured using noise feature extraction methods, resulted when tested at high levels of synthesized noise when compared to baseline performance. Noise feature extraction methods proved optimal and consistent, using a coefficient of 0.33, over the range of synthesized noise of 90-105 dBA. Speech input at or below 0 dB signal to noise ratio (same speech and noise audio level) produced optimum results, most often, when speech capture threshold levels were optimized. Speech capture threshold level adjustment is provided in the SCAD to preclude false trigger and to determine the on-set or end-point speech.



Testing of the SCAD in the presence of high noise level battlefield sounds showed superior speech capture characteristics when using the optimized noise feature subtraction coefficients. Pre-Recorded sounds of tactical vehicles in a battlefield environment are used to benchmark the performance of the SCAD in a realistic complex acoustic spectrum. Research testing was successfully conducted at approximately 105 dBA noise levels for a variety of tactical characteristic sounds with significant improvement in speech recognition accuracy scores and vocabulary word distinction. It is noted that the improvement in speech capture accuracy derived from stationary noise feature extraction techniques is in addition to the already defined baseline. The baseline already has significant performance improvements inherent in the noise canceling microphone and the statistical random noise elimination algorithms. The baseline configuration permits the speech recognizer to continue operation in noise environments above 90 dBA to 100 dBA. The addition of the stationary noise feature extraction techniques extends operations up to 105 dBA and also improves accuracy by 15%, or more, above 90 dBA.

The Phase II development research successfully achieved the objective refinement and design for the miniaturized speech capture system for DoD and commercial applications in high noise environments.

Performance sensitivity testing was performed for varying noise levels and characterized sounds to define the algorithm coefficients and to refine the final design for a superior noise characterization and speech capture system. This SCAD prototype embodies the successful results of this research in a miniaturized and refined design suitable for portable use by the soldier. The speech capture system represents the leading edge technology in speech capture under high noise battlefield conditions and will result in a technologically advanced and cost-effective commercial product.

The goal was to achieve automated sequential noise characterization, noise feature generation, and noise feature subtraction by the SCAD without the need for operator intervention. This process required that the DSP, Flash PROM, and FPGA on the SCAD multichip module dynamically reconfigure itself given the detection of the speech on-set. The algorithms are complete with the insert of optimized coefficients and threshold values. These algorithms have been tested and executed on the DSP automating these sequential functions, without compromising the accuracy and real-time performance attributes.

#### **4.0 TECHNICAL DISCUSSION**

The developed Noise Feature Set Array subtraction method for improving speech capture was successfully demonstrated in 90 dBA, 100 dBA, and 105 dBA noise environments. It was determined that a set of consistent coefficients for feature subtraction algorithms proved optimal under all tested conditions of variable frequency, dB level and signal to noise ratio. Similar improvements in speech capture accuracy were recorded for individual stationary noise frequencies as well as complex harmonic rich noise environments containing tactical battlefield platforms. Stochastic cancellation of random noise continued to perform as expected in the

presence of stationary noise feature subtraction algorithms. The developed SCAD simultaneously processes algorithms to improve the capture of speech in very high level noise environments containing both random and stationary noise.

Synthesized fixed frequency stationary noise at 100 dBA and 105 dBA was used to determine optimum algorithm coefficients to expand the previous testing conducted at 90 dBA. Baseline speech capture performance was compared with noise feature subtraction methods to show improvement achievable at various background noise levels. Speech input level was researched as an independent variable and was examined in terms of signal to noise ratio to judge its effect. Optimum speech capture resulted when the signal was approximately at the same audio level or less than the noise under most conditions, which reinforces the knowledge that very loud speech is distorted.

These findings represent a breakthrough in speech processing and especially in speech recognition, that previously required a 5 to 10 dB level of speech signal higher than the noise for speech intelligibility. The noise feature extraction developed algorithms, when applied to the baseline speech recognition capability, continued to perform well at equal levels of speech and noise audio levels. Particular words in the speech recognition vocabulary, that did not contain plosive sounds, did not fare as well at the 105 dBA noise level.

Improved speech capture was demonstrated by the synthesized stationary frequency testing at 500 Hz, 1 KHz and 2 KHz. More improvement was noted at the higher frequencies due to the physical and electrical filter break points that are inherent in the speech capture hardware and channel components below 1 KHz. Testing above 2 KHz was not conducted since greater improvement in performance was predicted by the empirical data. Successful testing in the presence of M1A1 Tank field recorded complex spectral noise with range 50 Hz to 4 KHz at 100 dBA proved this assumption correct. Speech capture improvement in the presence of complex M1A1 noise was noted during testing to be similar to the average response data observed for the synthesized range of frequencies. Speech capture accuracy and word discernability improved 10% in the 100 dBA M1A1 noise environment and 15% on average in the presence of M113, Jet Aircraft and Helicopter noise.

#### **4.1 Noise Feature Set Subtraction Model and Algorithm Development**

##### *SCAD Application Software and Program Interface*

A development objective was to complete the application software design and implement the Noise Feature Set Array Subtraction Algorithm for the SCAD. The prototype SCAD PCMCIA card must incorporate the TERI developed APIs needed to capture, characterize and filter-out stationary noise from a sampled utterance. This utterance would then be matched against a stored vocabulary "on-board" in the SCAD memory, which was captured in a benign environment. Characterized and filtered utterances produce

higher confidence recognition scores against the stored vocabulary as compared to unfiltered utterances.

Software application development began by defining an approach by which the prototype SCAD card could perform the recognition tasks. Figure 1, "SCAD Application Flow Diagram," depicts this approach. The flow diagram's process and decision nodes are numbered and described in Table 1, "SCAD Application Design." Application software development was written in ANSI "C" language and its compiled output is a DOS/Windows compliant .EXE executable file.

In order to properly ascertain the functionality of the SCAD Feature Set Array subtraction algorithms, trial iterations under varying noise conditions were tested. The concept was to characterize the environmental noise, then manually choose what type of noise to subtract from the Signal plus Noise Feature Set Array. This hypothesis is then tested through experimentation.

The tests conducted required a modification in the application software to allow the user to manually select a Noise Feature Set Array (500 Hz, 1 KHz or 2 KHz) to be subtracted. Signal plus Noise Feature Set Arrays were then captured under varying conditions (No Noise, 500 Hz, 1 KHz and 2 KHz). If there was no noise in the environment, and we subtracted Noise features from the Signal plus Noise Feature Set Array, the recognition scores would get worse (go higher). This is because instead of subtracting noise we were actually subduing features of the signal itself. Likewise, if we injected 500 Hz noise into the environment but subtracted out features of 1 KHz noise, the scores would also increase. The only trials that would produce better (lower) recognition scores is when the environment noise matched the Noise Feature Set Array that was subtracted.

The application software necessary to capture, characterize and analyze a Signal plus Noise utterance, then subsequently subtract stationary noise features and compare it to a stored vocabulary, has been completed. A number of "C" language modules that perform the recognition, memory allocation, Feature Set Array subtraction and API incorporation were created. The modules are compiled together under a single project file whose output is provided as a standard form [.EXE executable file.]

The procedures and results to test varying stationary noise frequencies (none, 500 Hz, 1 KHz and 2 KHz) against the frequencies used from the stored Noise Feature Set Array for subtraction, are attached. A multiplier coefficient (range 0-1.0) was used to vary the intensity of the stored Noise Feature Set Array used for subtraction. A factor of 1.0 meant that the entire Noise Feature Set Array was used for subtraction, typically about 90 dBA. As the coefficient decreased to 0.2, the intensity of the Noise Features subtracted was reduced to a fraction of its original captured levels. A baseline test was also conducted before any Noise Feature Set Array Subtraction was done.

**TABLE 1: SCAD APPLICATION DESIGN**

(1) Open SCAD Hardware	Utilizing the <i>Open</i> API, software controls the initialization of the hardware which includes IRQ and Base Address identification.
(2) Is Board Open?	If IRQ and Base Address conflicts occur an Error Message is processed and the demonstration is exited, else continue.
(3) Initialize Vocabulary	Structures and data pointers allocate on-board memory to store the vocabulary. Utilizing on-board memory (access time 15 ns) versus PC memory (70 ns) speeds up recognition and initialization tasks.
(4) Initialization Complete?	If initialization of structures and data pointers was not completed an error message is generated and the demonstration is exited, else continue.
(5) Open Vocabulary	Using the <i>Open Voc</i> API, the SCAD hardware is loaded with a user defined vocabulary. Data Pointers and Function Structures are allocated.
(6) Vocabulary Opened?	If the desired vocabulary is not opened properly, an error message is generated and the demonstration is exited, else continue.
(7) Load Global Parameters	Parameter files are associated with each vocabulary individually. They provide the user with definable parameters which affect the performance of the SCAD. These parameters include threshold audio levels, recognition acceptance score levels and vocabulary training ceilings.

**TABLE 1: SCAD APPLICATION DESIGN (CONTINUED)**

(8) Get Response	Captures a Signal Plus Noise utterance using the <i>Get Speech Utterance</i> API. Signal processing occurs to filter out non-stationary noise, time normalize the utterance then match its Feature Set Array to those stored in the on-board vocabulary.
(9) Response OK?	If an utterance response error is generated (e.g., spoke to soon or spoke to low) the user is returned to Get Response until a valid utterance is captured.
(10) Display Recognition Results	Results of the matched Feature Set Array capture (Signal plus Noise) to those stored on-board (Signal only) are displayed. Results reveal confidence and discernibility of desired match to all words in the vocabulary. Confidence level is given by a score between 1-1999, the lower the better, and discernibility is determined by the difference (or delta score) between the first and second choice given, the higher the better. The top four choices are displayed.
(11) Capture Noise	Captures Noise only utterance. Will be used to subtract stationary noise from Signal Plus Noise Feature Set Array.
(12) Convert Utterance	Converts Noise only utterance to Feature Set Array to match the structure of the Signal plus Noise Feature Set Array.

**TABLE 1: SCAD APPLICATION DESIGN - (CONCLUDED)**

(13) Perform Feature Set Array Subtraction	Subtracts out intruding Noise features from Signal plus Noise Feature Set Array.
(14) Create New Utterance	Product of Feature Set Array subtractions converted into standardized utterance format to be passed back to the recognizer for reevaluation.
(15) Perform Recognition Match on New Utterance	Feature Set Array match of new Signal plus Noise minus Noise utterance to stored vocabulary words is completed.
(16) Display Recognition Results	Confidence and discernability scores are regenerated and displayed for new Feature Set Array match.
(17) Deallocate and Free Memory	Uses both <i>Close Voc</i> and <i>Close APIs</i> to deallocate memory and free data pointers and structures.

Analysis of the experiments have shown the concept of noise feature subtraction built into the SCAD prototype performs as expected. Signals (speech utterances) captured in a benign environment (no Noise), got worse after noise feature subtraction. This is because subtracting out features from the Signal plus Noise Feature Set Array that are purely related to Signal are detrimental. Similarly, other trials in which the recognition scores worsened (go higher), were when the wrong noise features are used in subtraction. It is logical that if a 1 KHz noise frequency existed in the environment and the features associated with 500 Hz frequency noise were removed that the Signal plus Noise utterance recognition would get worse. The only trials that showed improved recognition accuracy was achieved when the environment stationary noise exactly matched the subtracted Noise Feature Set Array.

It was noted that the best recognition improvement was achieved when the subtraction multiplier coefficient was in the range of 0.2 to 0.33. Both the Signal and Noise Features reside in the same frequency bins of their respective Feature Set Arrays. It is therefore not prudent to completely subtract all the Noise Features (coefficient=1) from the Signal plus Noise Feature Set Array. When the amount of noise being subtracted was scaled, leaving some of the power remaining in those Signal plus Noise frequency bins of the array, the recognition accuracy improved.

Current implementation of the Noise Feature Set Array subtraction task is accomplished using noise templates stored on-board in memory. Ultimately, the SCAD hardware/software must work together in "real-time" to constantly sample true environmental noise, then just prior to the onset of a user requested *Get Speech Utterance* API (by pressing the Push-to-Talk switch), the latest noise sample would be stored in memory and then used to filter (Subtract) the intruding noise from the Signal plus Noise utterance. Future research effort needs to be focused on the refinement of both the hardware and software to complete this task.

#### *Algorithm Refinement and Performance Testing*

Previous research indicated performance variation in the quality of speech recognition accuracy that was dependent upon the amount of characterized noise which was subtracted from the captured utterance. It was determined that a direct subtraction of the Noise Feature Set Matrix from the Utterance Feature Set Matrix did not always result in an optimum recognition, even though improvements were evident. This phenomena is caused by the nature of speaker independent algorithms that normalize amplitude and time warp captured speech (or noise). It was decided that the introduction of a multiplier coefficient (range 0-1.0) would be applied to the Noise Matrix before Feature Set subtraction to determine performance sensitivity. This will not disrupt the efficient and already proven capture algorithms. Previous experiments have shown that speech recognition scores and delta confidence improvements are dependent upon the multiplier coefficient used.

A testing protocol was established to generate data on performance variation as a function of the multiplier coefficient. The research experiments anticipated that the coefficient choice may be a dependent variable and are dependent upon speech amplitude, noise amplitude, noise frequency, vocabulary choice and other independent variables. This data will be gathered and plotted to ascertain optimum coefficients for use in the final SCAD algorithm software. A large amount of testing research is envisioned because of the number of independent variables and expected number of degrees of freedom that must be statistically controlled during the experiments.

## **4.2 SCAD Hardware Architecture**

Prototype SCAD hardware has been architected as a low power multi-chip module laminate (MCM-L) incorporated in a Type II PCMCIA form factor. The miniaturized form factor provides a portable and PC compatible solution for capturing and analyzing audio samples. The PCMCIA card MCM has been built to include integrated chip A/D and D/A conversion for a full 16-bit accumulator (96 dB dynamic range equivalent) voiceband analog interface to process the audio input and output, enough Flash RAM to store captured speech and noise Feature Set Arrays, and an embedded Analog Devices 21ADSP55a Mixed DSP for real-time FFT computation of characterized speech and noise samples.

The prototype SCAD conforms to the IEEE PCMCIA Version 2.0 standards for PC hardware and software interface requirements. User interface microphone and speaker connections are provided through standard audio DIN connectors. On-board programmable signal level, impedance matching and bandpass filters are capable of interfacing with a wide range of commercial and military I/O devices, such as; dynamic microphones (typical output range 0.1 to 50.0 mV), noise-canceling microphones (typical output range 1.0 to 5.0 mV), headsets and speakers (4-16 ohms). Through the use of an external Push-to-Talk (PTT) switch, the user can control the on-set of a speech utterance to differentiate it from the automatic noise sampling and characterization mode of operation.

Making use of a flexible mixed DSP, the SCAD hardware will be able to mature as improvements to the speech processing algorithms evolve with field testing. On-chip Random Access Memory (RAM), which is external EPROM programmable, allows for easy code revision at low production costs without a commitment to one-time programmable Read Only Memory (ROM) devices. EPROM loading also supports dynamic loading and reconfiguration during operation which will allow for the real-time field modification of signal plus noise Feature Set Arrays in host configurations.

#### **4.3 Research Trials and Experiments**

To conduct the aforementioned SCAD prototype experiments, the following controlled conditions and set-up was established. A desktop PC running with a DOS/Windows operating system hosted the SCAD prototype PCMCIA card. Audio input to the SCAD card used a directional noise-canceling microphone headset assembly with handheld Push-to-Talk switch. Two (2) power amplified speakers, connected to a Kenwood Oscillating Frequency Generator, surrounded the tester providing the noise environment.

The procedures followed to conduct a test iteration were as follows. The user-tester sat at the desk with the PC on it and wore the headset. The microphone headset assembly was positioned so that the microphone itself was 1" away from the corner of the users mouth (this is standard microphone placement to avoid any false triggers produced by puffs of air directly in front of the users mouth). A background noise frequency (500 Hz, 1 KHz or 2 KHz) was selected on the Frequency Generator and the volume was adjusted to read 90 dBA (at the microphone) using a Realistic Sound Level meter. The noise frequency type and level were then recorded on the TERI SCAD Test Data Sheets.

The SCAD application program was then started. The user manually selected a frequency type and level, from a compiled list of pre-recorded noise frequencies that was held in memory, to match the background frequency and level desired for the test. A prompt from the application program would then ask the user to speak one of the words from the vocabulary already stored in its on-board memory, see Table 2, "SCAD Test Vocabulary." The vocabulary word chosen and the level at which it was spoken (80, 90 or 100 dBA) was recorded on the test data sheet. Different audio input levels were tested to determine if spoken utterance variations would affect



the recognition quality. Different input levels would change the amount of power stored within particular frequency bins of the captured utterance Feature Set array.

**TABLE 2: SCAD TEST VOCABULARY**

zero	four	eight	set-up	exit
one	five	niner	low	backup
two	six	alerts	medium	delete
three	seven	reports	high	return

### *90 dBA Synthesized Noise Trials*

Tests were conducted on each vocabulary word (20), at each spoken utterance input level (80, 90 and 100 dBA), for each background noise frequency (500 Hz, 1 KHz and 2 KHz) at 90 dBA. After the desired utterance had been captured, the application controlled SCAD returns a recognition confidence score and delta. This information is used as a baseline to ascertain the quality of the recognition in terms of confidence scores as a function of background coefficient subtraction levels of 1.0, 0.33 and 0.2. The recognition confidence scores (lower is better) and confidence deltas (higher is better) were recorded on the test data sheets. The test data was then averaged for the entire vocabulary to ascertain performance sensitivity to the independent variables of noise and coefficient levels. Results are shown in Figure 2, "Confidence Scores with 90 dB Noise @ 500 Hz," and Figure 3, "Confidence Delta with 90 dB Noise @ 500 Hz." Similar tests were conducted at other noise frequencies as shown in Figure 4, "Confidence Scores with 90 dB Noise @ 1 KHz," and Figure 5, "Confidence Delta with 90 dB Noise @ 1 KHz," and Figure 6, "Confidence Scores with 90 dB Noise @ 2 KHz," and Figure 7, "Confidence Delta with 90 dB Noise @ 2 KHz."

The data presented in these six Figures confirm and expand the research results compiled originally. Test results confirm the hypothesis that performing recognition confidence scores and deltas (Signal Plus Noise) on stored vocabulary words with the aid of Noise Feature Set Array subtraction improves performance. The Baseline recognition performance data in Figures 2, 4 and 6 shows that recognition scores are higher (less desirable) before Noise Feature Set Array subtraction for the coefficient subtraction factors of 0.2 and 0.33. It has also been concluded that the coefficient subtraction factor 1.0 (subtraction of the entire Noise Feature Set Array), is least desired for all noise frequencies tested. This is caused by subtracting out too much power, from the frequency bins within those Feature Set Arrays, that match the desired signal utterance.

Various background noise frequencies (500 Hz, 1 KHz and 2 KHz) were used to determine how that variable would affect recognition confidence scores and deltas. Confidence scores improved (decreased) with coefficient subtraction factors of 0.2 and 0.33. Recognition

confidence deltas improve (increased), as compared with Baseline trials, when coefficient subtraction factors of 0.2 and 0.33 were implemented. Since the SCAD prototype card is inherently able to perform over a range of background noise frequencies, the recognition shows almost similar scores and deltas before and after Feature Set subtraction in some frequency bands. This SCAD inherent electronic component and dynamic filtering, when coupled with the low end frequency cut-off of the microphone, is most probably masking the test results.

An important aspect of the test research was to see what effect the audio level of the spoken utterance (80, 90 or 100 dBA) would have on the recognition confidence. Overall, the most desirable results were produced when the Noise Feature Set Array used for subtraction was scaled down by a factor of 0.33. It can also be concluded that even though the SCAD prototype can handle a wide range of spoken utterance input levels, a background noise level of 90 dBA and spoken utterance level of 90 dBA showed the most improvement at all frequencies.

Confidence scores for recognition accuracy improved most dramatically as the noise frequency was shifted to the higher end of the audio band ( $> 1$  Kz). Similarly, dramatic improvements in confidence delta scores were recorded under higher frequency noise conditions. This phenomena was probably caused by the "masking" effect of the SCAD built-in features that have a low frequency cut-off with a break point near 1 KHz. The noise canceling microphone was also contributory to this effect. Additional research experiments will be conducted with other input transducers to gain more insight on the effectiveness of the feature extraction method without the masking effects. It was also noted that additional coefficient selections between 0.33 and 1.00 may be necessary to find a truly optimum value, especially for high frequency stationary noise.

The Feature Set subtraction method indicated peak performance for most trials when the signal to noise ratio (S/N) was 0 dB (e.g., equal audio volume level). It is expected that recognition performance would be better when the speech level is much higher than the noise (greater than 0 db) under high noise conditions (e.g., 100<sup>+</sup> dBA). However, the Feature Set extraction method improved recognition under both the positive and negative signal to noise ratio conditions when noise was set at 90 dBA. Additional testing will be required at other levels of noise to determine the sensitivity of that independent variable on the judicious selection of the multiplier coefficient factor.

Data values for the gross optimum multiplier coefficient (0.33) were extracted from Figures 11 thru 16 to illustrate the relative improvements in recognition confidence scores and deltas for two key independent variables: Noise Frequency and S/N Ratio. Figure 8, "Confidence Score and Delta Improvement for Various Background Noise Frequencies," and Figure 9, "Confidence Score and Delta Improvement for Various Signal-to-Noise Ratios," give a measure of the effectiveness of the Feature Set noise extraction method as would be implemented in the SCAD at this time. Figure 8 indicated an average 18% improvement in captured speech recognition scores for a range of stationary noise frequencies using the Feature Extraction Method with coefficient of 0.33. Similarly, the confidence delta between the first and second

choice for the recognized word improved an average 10% at the lower frequencies and almost 100% at the high end (2 KHz). Figure 9 shows an average 17% improvement in captured speech recognition scores for a range of signal to noise ratios. The confidence delta also improved 27% on average, with the greatest improvement seen when the utterance audio was lower than the background noise. This phenomena holds great promise for the results of future testing at higher background noise levels. Additional research will be conducted at other noise levels and with protocols that eliminate component masking effects, to continue the research for determination of dynamic multiplier coefficients for overall performance improvement.

### *100 dBA Synthesized Noise Trials*

Tests were conducted on each of the 20 vocabulary words, at each spoken utterance input level (90, 100 and 105 dBA) with a background noise frequency of 500 Hz, 1 KHz, and 2 KHz each at 100 dBA. After the desired utterance had been captured, the application controlled SCAD returned a recognition confidence score and delta. With the increase in background noise from 90 to 100 dBA plus the added energy of 90, 100 and 105 dBA Signal, only about 50% of the 20 vocabulary words are properly recognized. Each of the other words, especially those with little starting plosive energy associated with them, such as eight or medium, could not be detected by the SCAD recognizer with that amount of background noise. The remainder of the properly recognized words were used as a baseline to ascertain the quality of the recognition in terms of confidence scores as a function of background coefficient subtraction levels of 1.0, 0.33 and 0.2. The recognition confidence scores (lower is better) and confidence deltas (higher is better) were recorded on the test data sheets. The test data was then averaged for the entire vocabulary to ascertain performance sensitivity to the independent variables of noise and coefficient levels.

Results are shown in:

- Figure 10, "Confidence Scores with 100 dB Noise @ 500 Hz,"
- Figure 11, "Confidence Delta with 100 dB Noise @ 500 Hz,"
- Figure 12, "Confidence Scores with 100 dB Noise @ 1 KHz,"
- Figure 13, "Confidence Delta with 100 dB Noise @ 1 KHz,"
- Figure 14, "Confidence Scores with 100 dB Noise @ 2 KHz,"
- and Figure 15, "Confidence Delta with 100 dB Noise @ 2 KHz."

Improved speech capture results are shown for each of the 100 dBA trails for both confidence score and delta values. The percent improvement in confidence score between baseline (no stationary noise reduction) trials and using a 0.33 noise reduction coefficient, in Figures 11, 13 and 15, indicated a 7%, 22% and 20% improvement for 500 Hz, 1 KHz and 2 KHz respectively. Confidence score recognition improvement was also seen at the 0.2 noise reduction coefficient values for each of the trials, but improvement was only about half of that obtained using the 0.33 subtraction coefficient. Full noise Feature Set Array Subtraction (using 1.0 subtraction coefficient) performed the same as the baseline without noise feature subtraction. This would indicate that removing too much noise and accompanying signal from the spoken

utterance Feature Set Array, is as damaging as leaving the intruding noise in the Feature Set Array in the first place. This result is true for both confidence score and delta trials.

An improvement in confidence delta was also achieved as seen in Figures 11, 13 and 15. The percentage improvement was 4%, 23% and 94% for 500 Hz, 1 KHz and 2 KHz respectively. When the improvement in confidence scores is compared to confidence deltas, it indicates that the SCAD recognition is more confident and produces a lower score when the delta between the recognized word and its next closest choice is high. Conversely, if the recognition confidence produces a higher score, the difference between it and the next closest choice would be lower and would exhibit a lower confidence delta.

Another result of interest exhibited by the graphic trial representation can be seen along the constant subtraction coefficients bars. Many take a "V" shape form being lower in the middle at 0 dB S/N and higher on the ends at both 0.9 dB and 1.05 dB S/N ratio. This data would lead to the conclusion that optimal performance will be obtained at a S/N ratio of 0 dB. At 0.9 dB and 1.05 dB S/N ratios the spoken utterance, volume and distortion respectively, are not balanced with the amount of Feature Set extraction giving sub-optimal results.

#### *100 dBA Field Noise Simulation Trials*

Now that the SCAD prototype card has been tested at various input levels for known single frequency background noise trails, the next task was to evaluate its performance using multiple frequency/pattern background noise sources. Those noise patterns were chosen from high quality digital noise recordings from the Army's Simulation Theater at Aberdeen Proving Ground (APG) in Aberdeen, MD. They included an M1A1 Main Battle Tank, an M113 Assault Vehicle, a Helicopter and a Jet. Tests were conducted using the same protocol and data collection methodology as the prior synthesized fixed frequency SCAD testing procedures. Results for the tests conducted using the M1A1 as the background noise @ 100 dBA is shown in Figure 16, "Confidence Scores with 100 dB M1A1 Background Noise," and Figure 17, "Confidence Delta with 100 dB M1A1 Background Noise."

Figure 16 shows that even though all the performance scores obtained were fairly close to one another, optimal performance was obtained at a noise feature subtraction ratio of 0.33 and at a signal to noise ratio of 0 dB (here shown as 100 dBA). A subtraction ratio of 0.33 at 0 dB S/N produces optimal recognition results for complex harmonic rich stationary noise environments.

These results are consistent with those obtained for the single frequency background noise optimization performance trials. The remaining field noise trial (M113, Helicopter and Jet) tests were conducted only at baseline (before noise reduction) and at the optimal 0.33 subtraction coefficient at 0 dB S/N (after noise reduction) results for 90 dBA signal and noise input levels. These results are shown in Figure 18, "Confidence Scores for Various Pre-Recorded Field Noise Trials" and Figure 19, "Confidence Delta for Various Pre-Recorded Field Noise Trials." What we can determine from these graphs is that for 6 out of the 8 tests (3 out of 4 in each of the

confidence score and delta results) recognition confidence improved with the use of the noise reduction techniques. The only test iterations that did not improve were the confidence score for the M113 (went higher) and confidence delta for the M1A1 (went lower). We must note however, that the SCAD continued to function at baseline levels and that the change was not dramatic (less than 5%) compared with the improvements made in the other 6 cases. SCAD performance dramatically improves when background stationary noise contains higher frequency components.

Figure 20, "Confidence Scores for Various Pre-Recorded Field Noise Trials," and Figure 21, "Confidence Delta for Various Pre-Recorded Field Noise Trials," show the results of the baseline and 0.33 subtraction coefficient tests at 0 dB S/N for 100 dBA signal and noise input level. Here, all 8 tests using the SCAD noise reduction schemes improved the scores (decreased) and deltas (increased). In most cases, the improvement in recognition confidence improved by better than 15%.

It is noted that the improvement in speech capture accuracy derived from stationary noise feature extraction techniques is in addition to the already defined baseline. The baseline already has significant performance improvements inherent in the noise canceling microphone and the statistical random noise elimination algorithms. The baseline configuration permits the speech recognizer to continue operation in noise environments above 90 dBA to 100 dBA. The addition of the stationary noise feature extraction techniques extends operations up to 105 dBA and also improves accuracy by 15% or more above 90 dBA, dependent upon the spectral content of the noise.

#### *105 dBA Synthesized Noise Trials*

Identical setup and test procedures were once again followed to collect the data for each background noise frequency now at 105 dBA.

The results are shown in:

- Figure 22, "Confidence Scores with 105 dB Noise @ 500 Hz,"
- Figure 23, "Confidence Delta with 105 dB Noise @ 500 Hz,"
- Figure 24, "Confidence Scores with 105 dB Noise @ 1 KHz,"
- Figure 25, "Confidence Delta with 105 dB Noise @ 1 KHz,"
- Figure 26, "Confidence Scores with 105 dB Noise @ 2 KHz,"
- Figure 27, "Confidence Delta with 105 dB Noise @ 2 KHz,"

Similar to the conclusions made in the 100 dBA background noise tests, a background noise of 105 dBA also shows marked improvement. The confidence score improvement seen in Figures 22, 24 and 26 for the 0.33 subtraction coefficient compared with the baseline (no stationary noise subtraction) delta is 15%, 20% and 21% for 500 Hz, 1 KHz and 2 KHz respectively. Optimal results were found using the 0.33 subtraction coefficient, however,

improvements were also made at the 0.2 subtraction coefficient across all trials. Once again, it is shown that the degradation to the Signal + Noise Feature Set Array at a subtraction coefficient of 1.0 produces scores close to those obtained in the baseline trials.

Figures 23, 25 and 27 show an even more dramatic speech capture improvement in confidence delta than in any other set of trials. The percent improvement seen compared to the baseline tests were 31%, 104% and 255% at 500 Hz, 1 KHz and 2 KHz respectively.

These dramatic improvements at 105 dBA verify the postulated performance predicted by the theory under investigation, that systematic removal of stationary noise energy Feature Sets would improve speech capture.

The "V" shaped bar graph data that we saw at 90 dBA and 100 dBA background noise levels are not present at 105 dBA. This is due to the amount of saturation in the Signal + Noise Feature Set Array at these high input energy levels. Optimal performance was obtained at a S/N ratio of 0.85 dB (at 90 dBA signal) at 105 dB. Improvements in speech captured were also seen at the 0.95 dB S/N trials (at 100 dB) at 105 dB. Improved (lower) recognition confidence scores and improved (higher) recognition confidence deltas were consistent for all subtraction coefficient values at a S/N ratio of 0.85 dB.

Results have shown both data sets to be consistent with those evaluated in prior iterations at 500 Hz, 1 KHz and 2 KHz with background noise at 90 dBA. Those results concluded that optimal SCAD performance was found at a subtraction coefficient of 0.33 where only 33% of the stored Noise Feature Set Array was used for subtraction from the captured S+N Feature Set Array. This is where we find the lowest scores (better) in Figures 11-19 (odd) and highest deltas in Figures 12-20 (even). Further investigation also supports the hypothesis that performance is optimized when the input signal level matches or is less than that of the Noise Feature Set Array being subtracted, in these cases 100 dBA and 105 dBA respectively. Even though, at these equivalent S/N Ratios (0 dB), performance is optimized for most subtraction coefficients, the best results are obtained using the 0.33 subtraction coefficient.

The increase in absolute scores and decrease in deltas at 100 and 105 dBA levels can be directly attributed to the amount of background noise. This could have been concluded even prior to testing but now has been verified. The SCAD prototype card has established its usability in noise levels of 105 dBA, with performance predicated upon low S/N Ratio. This is reasonable since it is unnatural to speak at levels above 90 dBA and in doing so, the user has altered the signal being presented to the SCAD card. The vocabulary stored on the SCAD was captured in a benign environment by users speaking in a normal tone, usually between 70 and 85 dBA. When the user essentially "yells" into the microphone, the spoken utterance patterns presented to the SCAD card for recognition do not match as well.

## 5.0 CONCLUSIONS

The result of the research is the quantification of both the anticipated noise and channel environments as well as the methodology and quantifiable means to demonstrate performance in that environment. The project started with a high quality transducer/signal processor and independent speaker recognition as a baseline capability derived from Phase I. In a structured and methodical way each of the noise immunity schemes including new algorithms and HW/SW techniques are brought to bear and documented to determine their usefulness and capability to enhance the speech capture quality in a noise environment. TERI has qualified speech capture and computer recognition system products and we have experience in each of the noise immunity techniques.

This research provided the foundation research in post processing random and stationary noise treatment that quantified the feasibility of the postulated methods. We proceeded with confidence to the Phase II development and testing which culminated with the demonstration in a tactical field simulated noise environment. The results of the research are provided in a description and report of the speech capture improvement experiments in response to the noise immunity software algorithms and ancillary devices.

Development of the SCAD has been productized to automate the capture of background noise in real-time and then apply that characterization as a remedy to improve speech capture in that instant noise environment. In order to accomplish this, low-level firmware assembly code had to be developed to support the continuous background noise capture sub-routine and functions within the limits of SCAD's on-board Program Memory (PM) space. These functions are now combined with the existing SCAD speech recognition and FFT routines.

Prior results from the SCAD's performance benchmarking tests had already concluded that the subtraction of both stationary and non-stationary background Noise Feature Set Arrays from captured Signal plus Noise Feature Set Arrays improved recognition confidence scores (goes lower) as well as confidence deltas (goes higher). The ability of the SCAD to properly recognize the spoken utterance in the presence of noise from a defined set of vocabulary words, and how well it can discern that utterance from other words within the vocabulary, ultimately measures the performance of the recognition engine. Optimal recognition performance was found at a Signal to Noise (S/N) ratio of 1.0, where there are equal amounts of both Signal and Noise, and when a noise subtraction coefficient of 0.33 is used. Noise subtraction coefficients were tested to determine what ratio (from 0.25 to 1.0) of noise was to be used for subtraction. We found that a direct subtraction of the Noise Feature Set Array from the Signal plus Noise Feature Array was too detrimental to the overall utterance because similar vocal energy would reside in like frequency bins between the two arrays.

Knowing this, our task was to create a standalone solution that would continuously monitor any background noise from 70 dBA to 105 dBA. The captured background noise would

then be stored on-board, in 1.5 second intervals, overwriting itself every cycle to save Data Memory (DM) space and also to ensure that what's captured on-board is the latest representation of the background noise in the users environment. When the user initiates an utterance through a press of the PTT, a signal plus utterance is captured. A Noise Feature Set Array coefficient would then be applied to last known good background noise capture and its result is subtracted from the signal plus noise utterance. Utterance recognition results, in terms of confidence score and delta, are visually presented by the application program to the user both before Noise Feature Set Array subtraction and after for comparison.

A working relationship between the host application and the SCAD firmware is supported by the application program which elicits the continuous noise capture function. Already defined SCAD API libraries handle the handshake and data transfer duties between the host computer and the SCADs firmware. The interface between host and SCAD use previously documented interfacing and memory-mapped structures. All internal SCAD data registers are accessed by the PCMCIA controller and have read/write capability directly with the host, without intervening external device drivers.

The successful development and integration of the API routines has produced a SCAD that can continuously monitor both stationary and non-stationary background noise between spoken utterances. The latest known good 1.5 second noise capture is repeatedly stored on-board in data memory awaiting the onset of a user initiated PTT. Visual feedback of the status for the background noise continuous capture routine signals the user that the SCAD detects background noise. Recognition confidence scores and deltas are displayed for user verification, both before and after the noise remedy techniques are applied. If no background noise is captured, either because the noise level was below 80 dBA (the minimum useful level), then no Noise Feature Set Array subtraction is performed and the recognition confidence scores and delta for the signal utterance only are displayed.

Tests were conducted with 5 spoken utterance words using a 20 word vocabulary. Tactical battlefield recorded background noise at 85, 90, 95, 100 and 105 dBA. The S/N ratio was approximately 1.0 for each trial. The averaged results are shown in Figure 28, "SCAD Confidence Score Comparison with Stationary Noise Feature Extraction," and Figure 29, "SCAD Confidence Delta Comparison with Stationary Noise Feature Extraction." The only variation from the initial SCAD development tests was the noise subtraction coefficient. Continued trials with the final SCAD product showed optimal performance was achieved at a Noise Feature Set subtraction coefficient of 0.125.

For the 10 tests conducted, one for confidence score and one for confidence delta at each of the 5 background noise levels, 9 out of 10 improved with noise reduction techniques applied (with the exception of confidence delta at 85 dBA), which means that the scores got lower and the deltas went higher. Even for the single low noise level trial that did not improve, the change was less than 1%. The utterance recognition improvement for all trials was about 10% on average for both confidence scores and deltas. Optimal performance was obtained at the lower



input signal and background noise levels of 85 and 90 dBA where we see the lowest scores and the highest deltas. This is directly associated with the normal levels of speech and also the level at which the stored vocabulary was created. At 95 dBA and above, the speaker begins to compete with the background noise and starts yelling. Spoken voice patterns become more distorted, and even though the SCAD was able to correctly identify the utterances (25 out of 25), the match that is made between the captured utterance and the stored vocabulary begins to show a progression in higher scores. Significant to the SCAD product test results is that the speech recognition and accurate utterance capture continues to perform without significant error up to and including 105 dBA levels of noise.

## **6.0 RESEARCH FUTURE APPLICATIONS**

The noise characterization benchmark methodology, which quantifies the performance of speech capture devices in a noise environment, and is a very valuable result of the research that is applicable to all future R&D effort in this field. It is expected that the benchmark methodology will provide a significant advantage for all future research in speech processing.

Significant outgrowth of the research is the productization of a device that will perform accurate speech capture in a noise environment, as a complete system. Research prior to this proposed effort as documented in the literature has been scattered in pockets of individual research with significant gaps between the research projects associated with speech capture and computer automation using voice commands. The approach researched will fill these gaps and bridge the various research projects to provide an integrated solution that is effective in the proposed man-machine interface environment. The controlled and quantified approach provides all future research in this field with a way to analyze performance and identify error source contributions.

The results of this topic research will serve as the building block for future research in the field of speech capture for automated processing. Development of benchmark methods for the assessment of speech capture systems is a significant contribution to this area of research. Using a digital signal processor to automatically quantify speech capture accuracy is innovative to this field which today is prone to subjective and manual assessment methods. The ability of the digital signal processor to accomplish the speech capture assessment, in real-time, is significant to future research that will apply remedies to counter noise conditions. The robust nature of the proposed methods will be influential to all speech capture research in either acoustic, thermal or electrically generated noise that would mask the speech signal.

The proposed research will also establish a "high water mark" for the performance of a speech capture system in a high noise environment. This will indicate the current status of the technology in this field of acoustics related to speech automated processing.

The application of the proposed project results has potential for use by the military and commercial applications in many areas. Speech recognition has proven to be a successful and

efficient method of human-machine interface especially in a complex task and high stress environment. The use of quality speech capture and voice recognition to support the military mission and to improve the performance of the man-in-the-loop has many applications to support crew members in aircraft and in vehicles as well as on board ships, in warehouses, dockside, on the battlefield for the individual soldier, in a High-Mobility Multi-purpose Wheeled Vehicle (HMMWV), and elsewhere. The tactical environment demands quality speech and voice recognition in many instances where the hands and eyes of the military crew or soldier may be preoccupied with other tasks thus giving him the additional option to interact with his computerized equipment utilizing voice commands. Typical applications for soldiers include the ability to interface with computers and communications equipment in an environment which requires the wearing of protective gear, gloves and masks which would preclude use of the traditional I/O devices. Extensions of the soldiers or crew member ability to continue to interface with his equipment in darkness as well as in severe weather and noise conditions represents other opportunities for military application to extend his efficiency and ensure successful completion of the assigned missions. The development of enhancements to state-of-the-art voice interaction capability will support the tactical mission in these extended environments and under noise conditions.

A speech capture system that can operate in a noise environment has many commercial applications as well as military. The basic research that will be developed as part of the benchmark capability and the data gathered for each of the HW/SW techniques for computer voice recognition in a noise environment applies to speech intelligibility in general. The art of speech intelligibility is a prime concern in all military and commercial applications be it telephone, radio, communications with computers or even providing hearing aid enhancements for the hearing impaired. The fundamental research that will be accomplished in speech intelligibility by humans as an interface to machines will help provide a quantifiable way to measure performance and speech intelligibility in a structured and repeatable way. Current literature admits to the gap in speech intelligibility as related to measurable signal-to-noise-ratios. The human ability to "track" speech and discern it in a noise environment will be embodied in the features and techniques which will be investigated in the form of software algorithms and speech segregation transducer electronics. The use of closed end vocabulary sets as well as knowledge based precognitive algorithms will reflect the human ability to anticipate speech input and therefore process and help separate it from the acoustic and channel noise background. The commercial application for the speech capture transducer and signal processor in a noise environment will include direct application to the manufacturing floor, robotics, noisy dock side activity, fire fighting, radio communications, in fact any area where man must communicate with machine in order to support his current efforts. Speech as a man-machine interface is becoming more and more important in every day activity as the electronics and communications permit miniaturization and portability where none had existed before. The interface between the man and the machine therefore becomes very important to include the use of voice commands in hands and eyes busy activities. Commercial applications also include the military parallel for use over remote channels such as in vehicles and aircraft which are acoustically and electronically noisy environments.

Anticipated use of the basic research also includes the advancement in separation of desirable signals from noise clutter in recorded and transmitted channelized audio information. TERI fully expects that the research and development work carried out here will have application in acoustic exploration and other sensor identification activity as a means to separate needed data from background clutter and either electronic and acoustic signatures. Other applications include voice controlled computers, typewriters security systems, communications systems, safety devices, inventory controls, robotic control and a host of other abilities associated with voice command and speaker identification which will be enhanced through this research and development. As technology for voice control systems becomes more advanced such systems will invade nearly every aspect of daily life since voice is the most natural man-machine interface.

Commercial application in the near term have been identified and dialoged with several organizations which are in process. Discussions have been held with MIDAS Muffler Co. for speech input of customer service and inventory/warranty into an automated database. Large banking institutions have shown interest in a speech capture system that would operate in a noisy lobby, casino, airport and shopping mall floor for access to the Automated Teller Machines. The provision of a superior speech capture system that has been miniaturized, such as proposed in this advanced development project, has wide commercial application since no prior training is required by the user. The general populace speech variation, accents and anomalies will all be acceptable to the proposed speaker independent speech capture microchip. The small size and relatively inexpensive implementation will make the speech capture system attractive for incorporation in household appliances, tools and portable convenience accessories that can use voice commands for control or selection. Vending machines, shopping dispenser kiosks and entertainment virtual reality games of the future will all be voice activated.

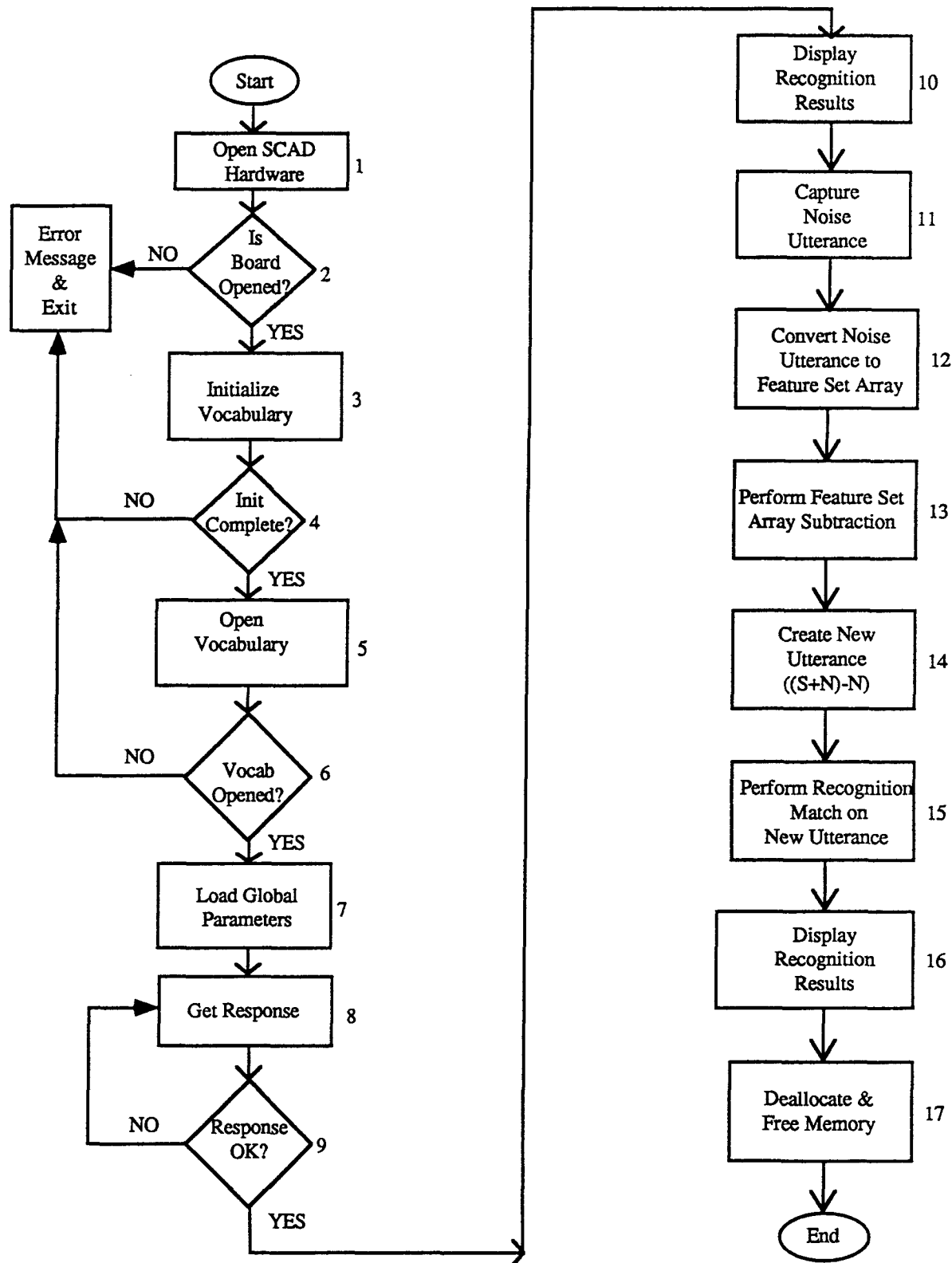
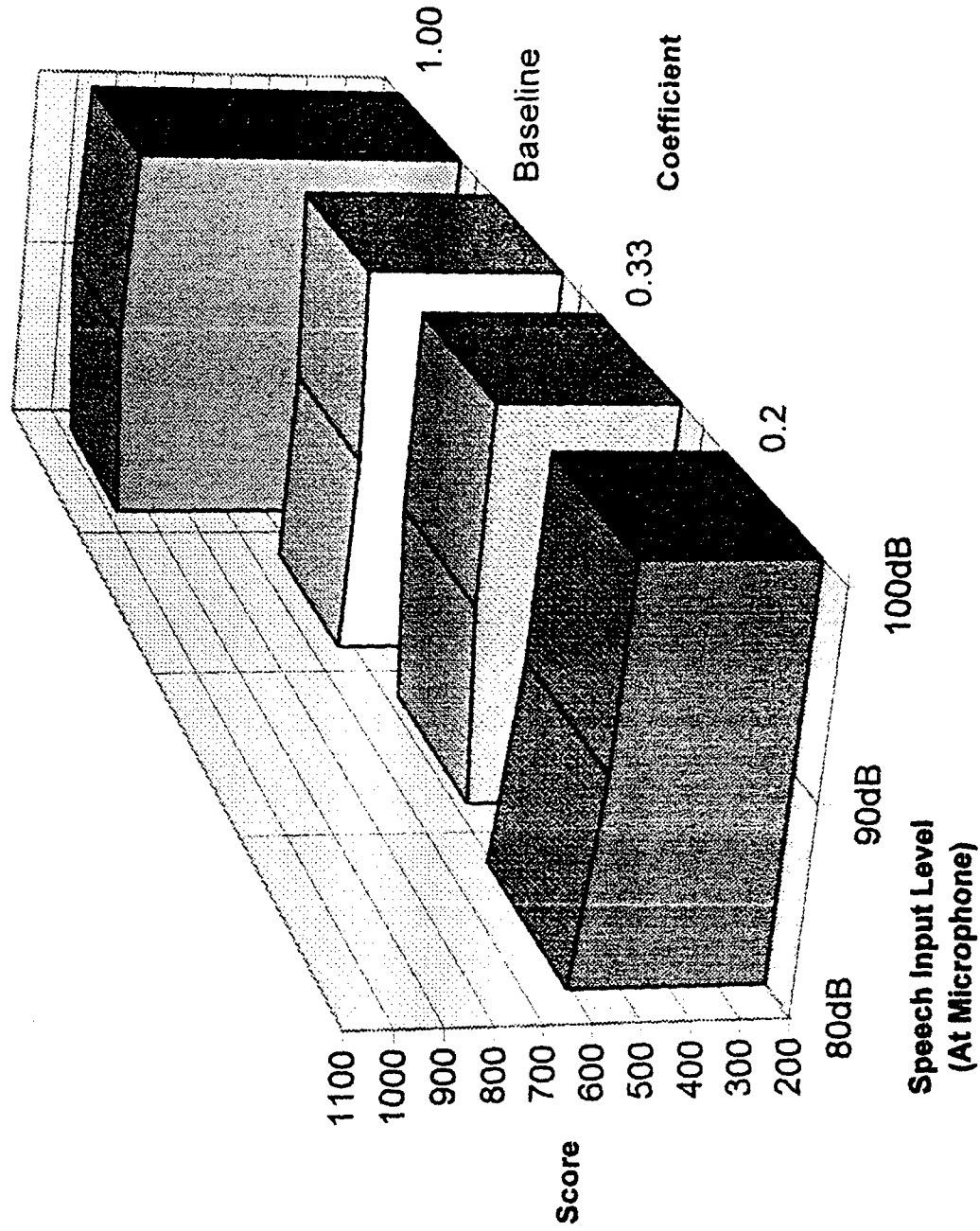


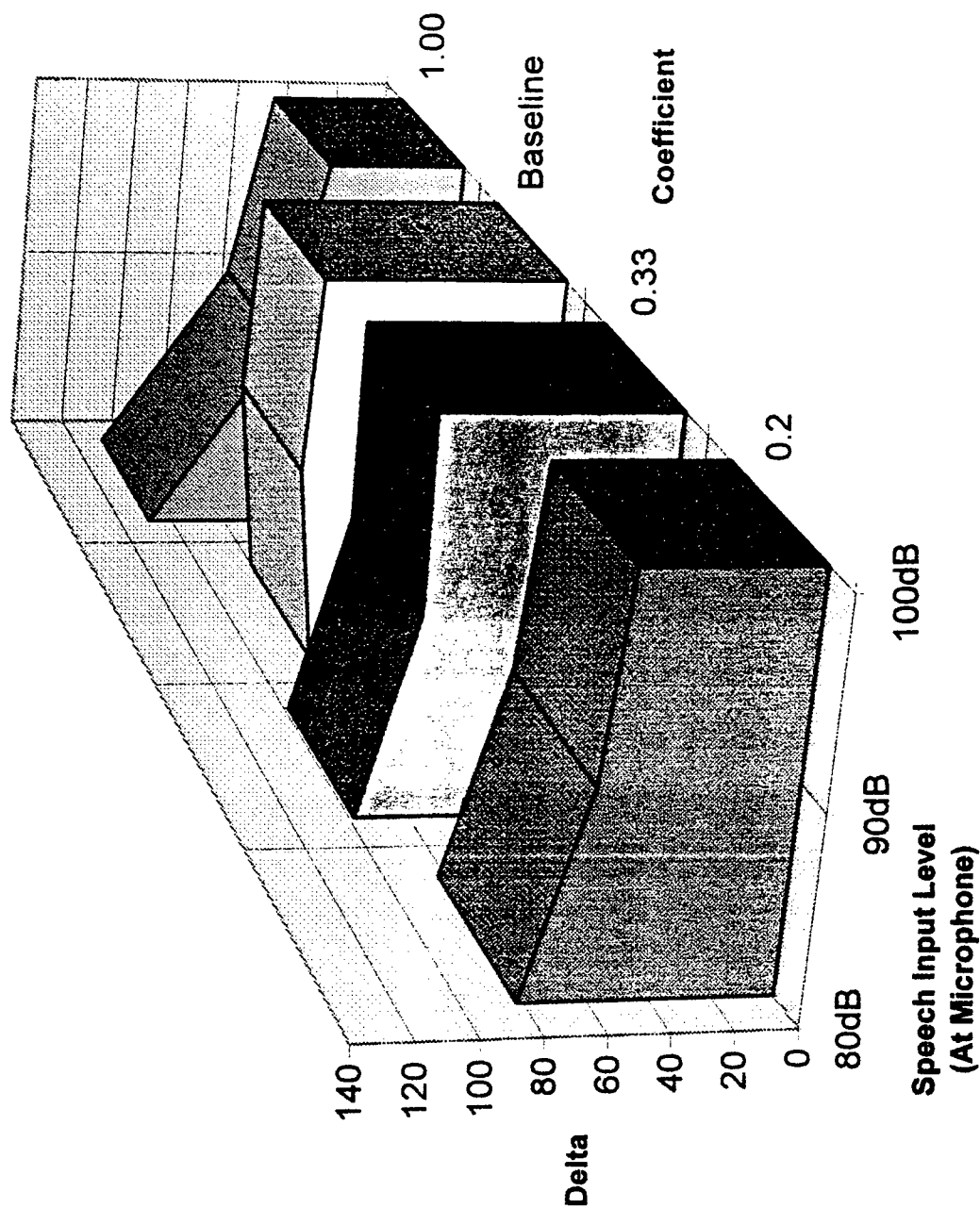
FIGURE 1: SCAD APPLICATION FLOW DIAGRAM

# **Confidence Scores vs. Noise Subtraction Coefficient For Various Speech Input Levels**



**FIGURE 2: CONFIDENCE SCORES WITH 90 dB NOISE @ 500 Hz**

### Confidence Delta vs. Noise Subtraction Coefficient For Various Speech Input Levels



**FIGURE 3: CONFIDENCE DELTA WITH 90 dB NOISE @ 500 Hz**

# Confidence Scores vs. Noise Subtraction Coefficient For Various Speech Input Levels

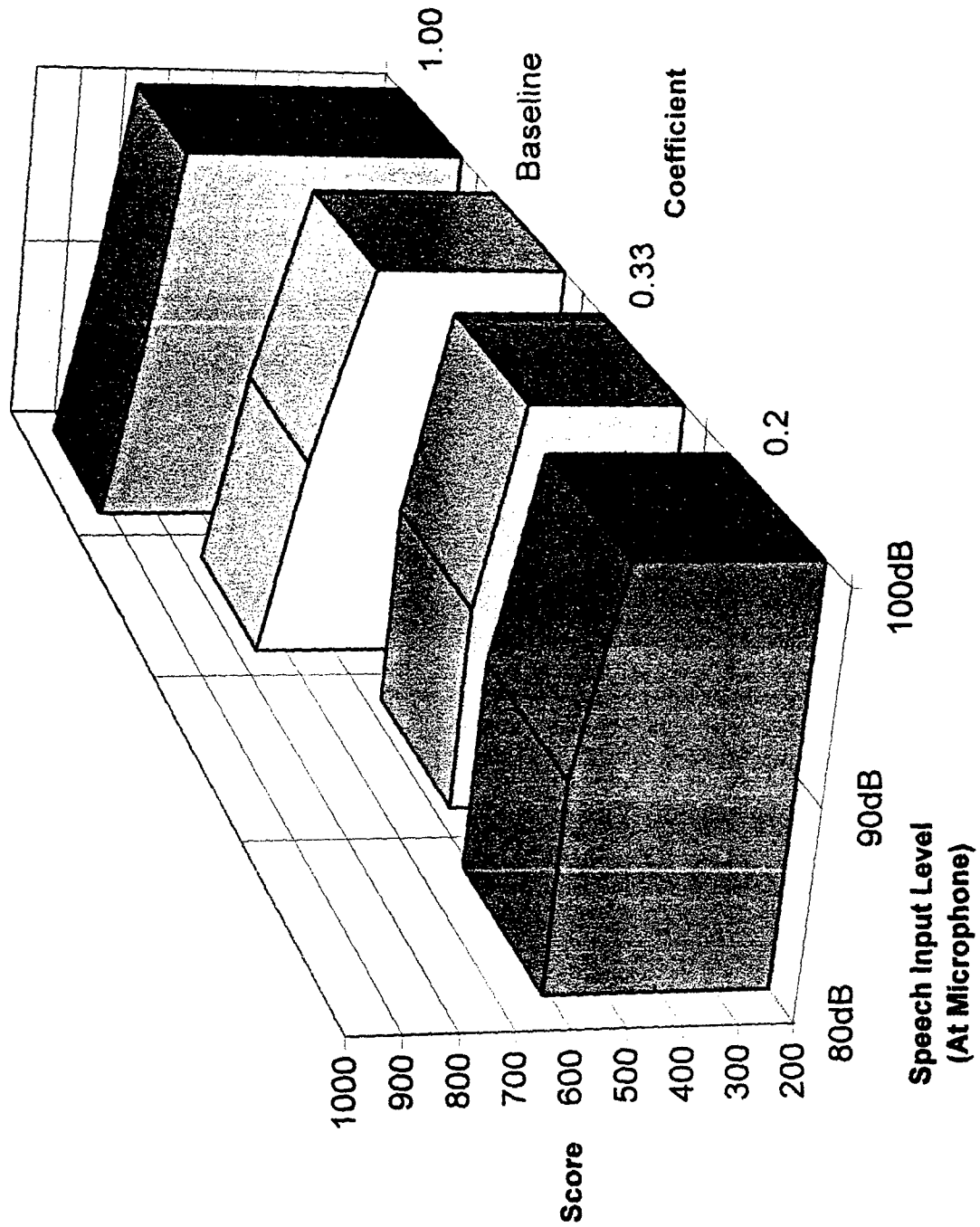
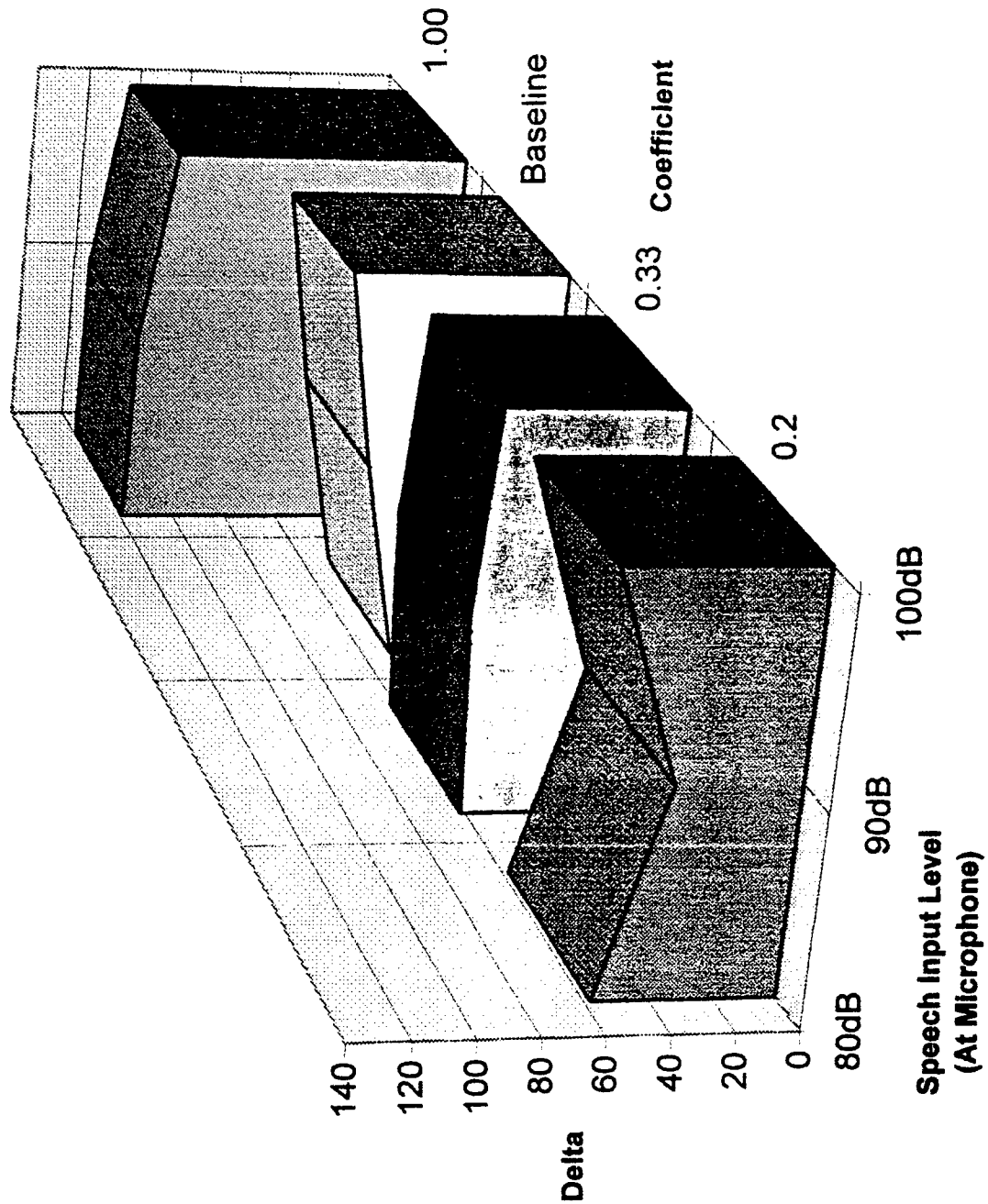


FIGURE 4: CONFIDENCE SCORES WITH 90 dB NOISE @ 1 KHz

# **Confidence Delta vs. Noise Subtraction Coefficient For Various Speech Input Levels**



**FIGURE 5: CONFIDENCE DELTA WITH 90 dB NOISE @ 1 KHz**



# Confidence Scores vs. Noise Subtraction Coefficient For Various Speech Input Levels

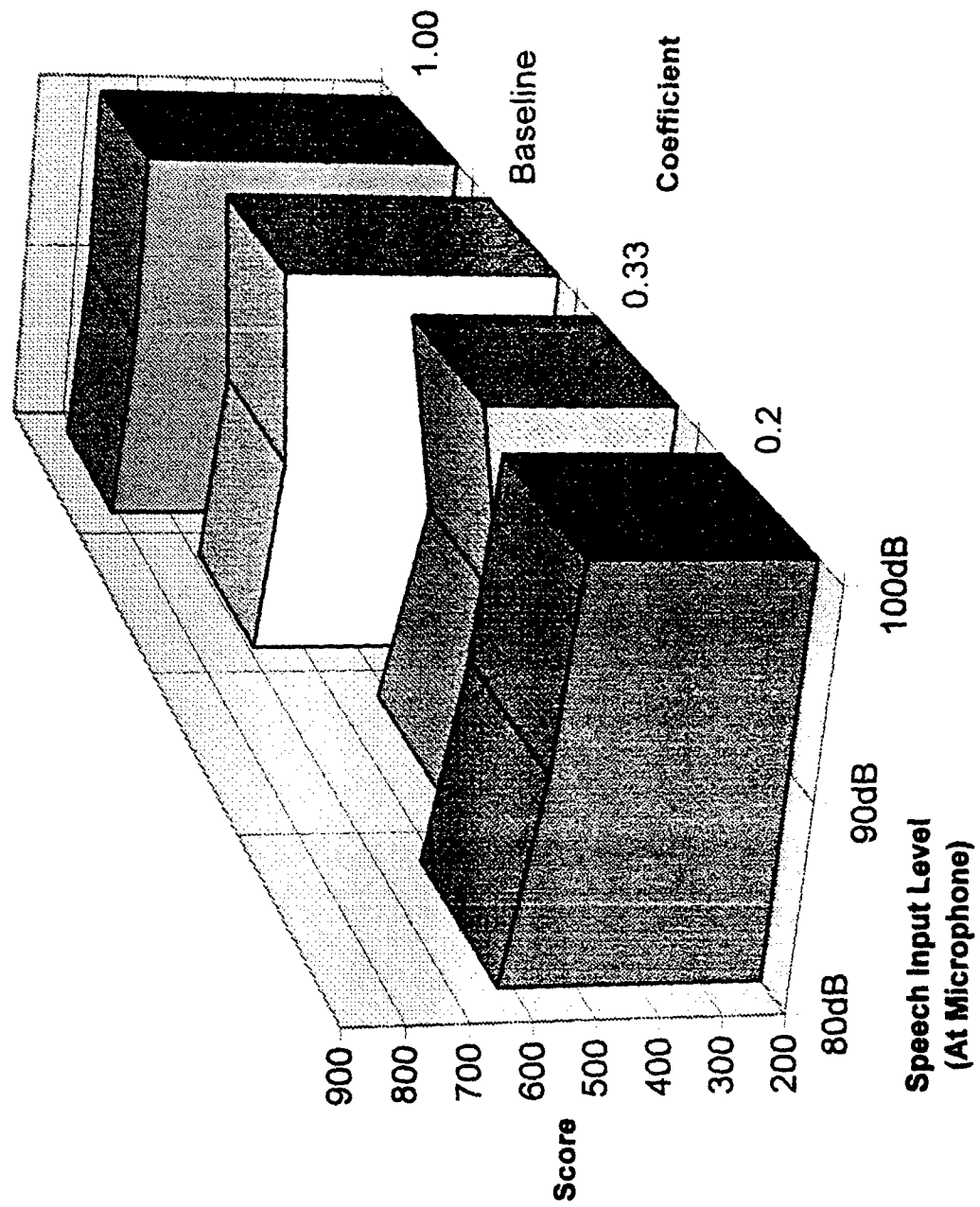
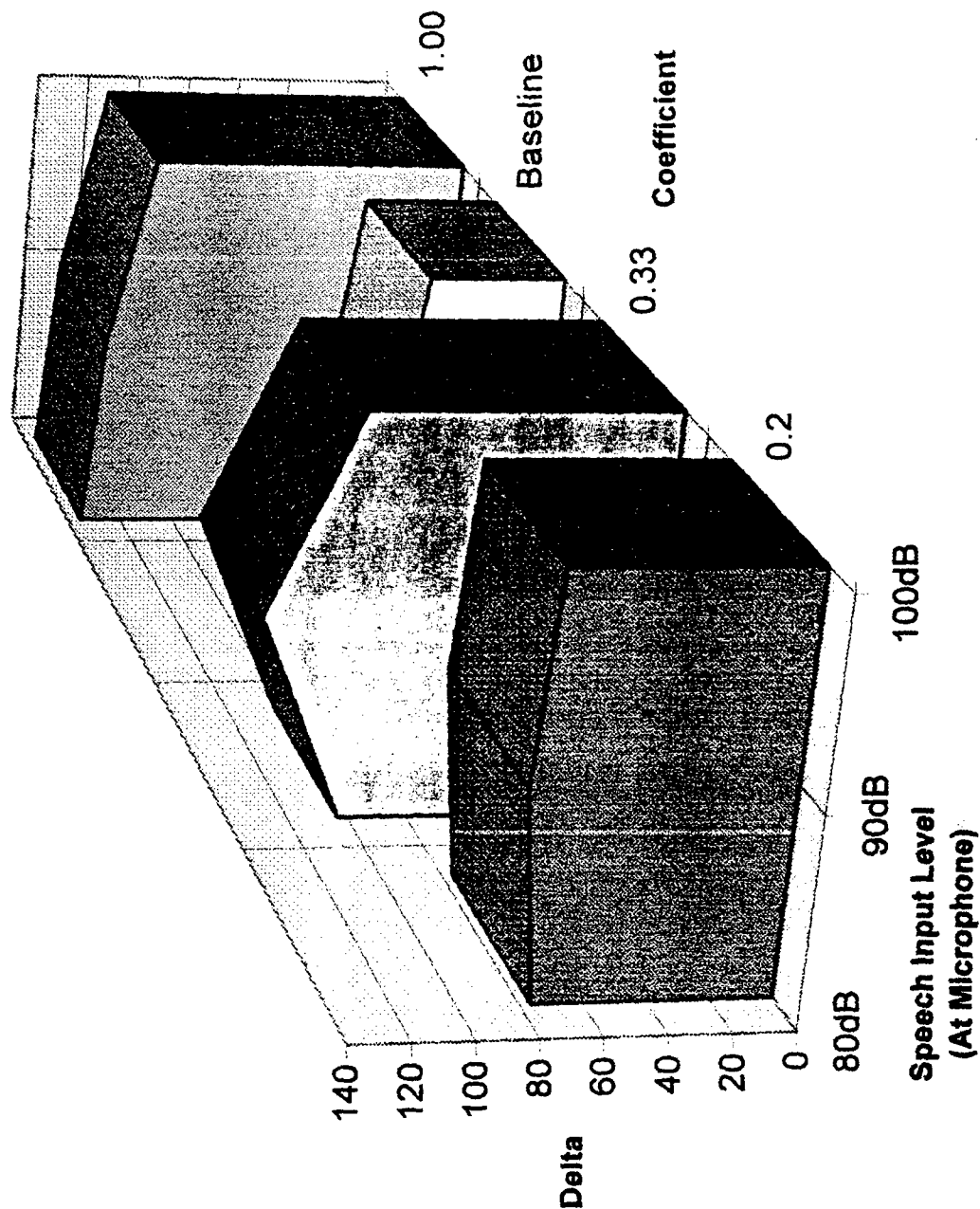
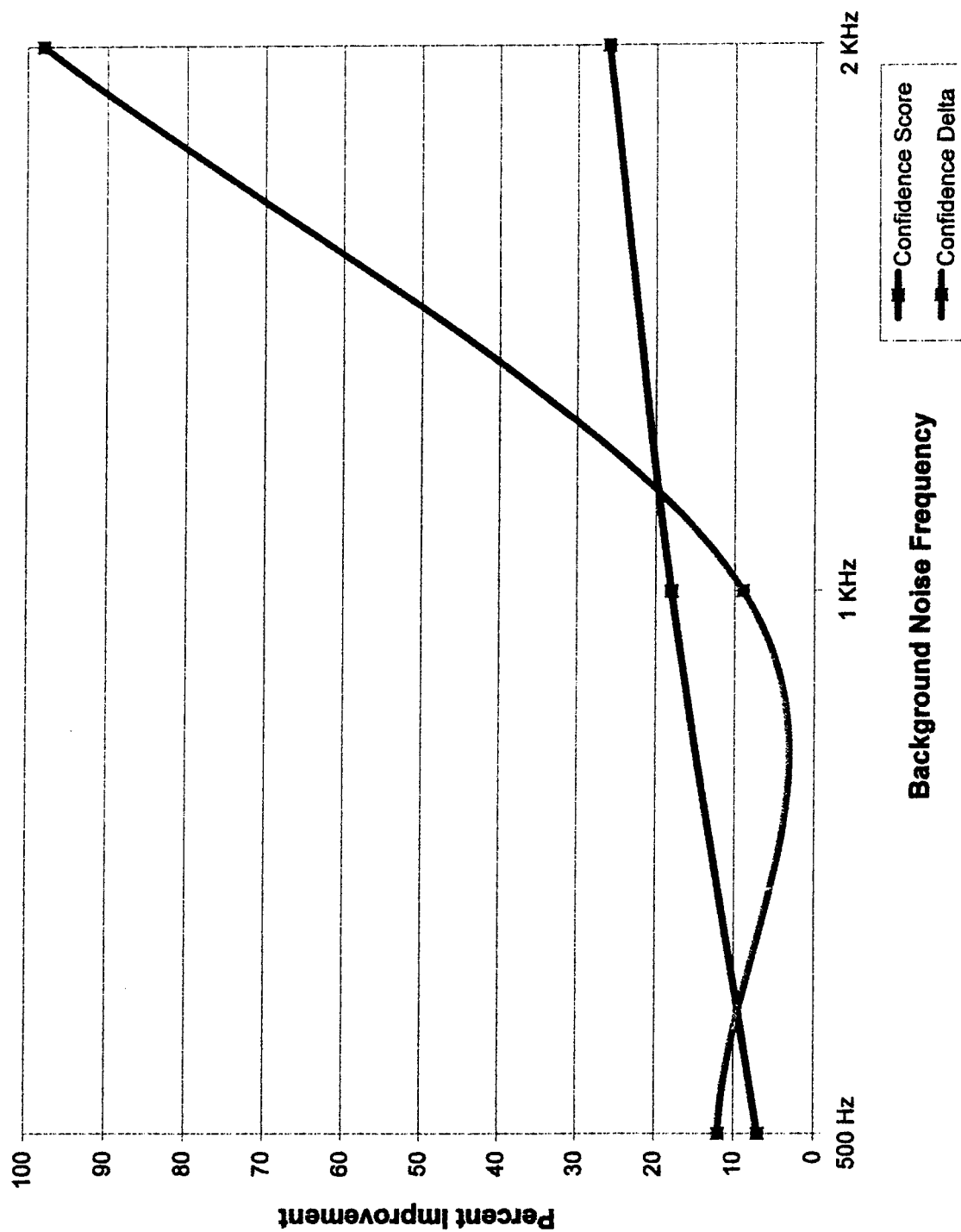


FIGURE 6: CONFIDENCE SCORES WITH 90 dB NOISE @ 2 KHz

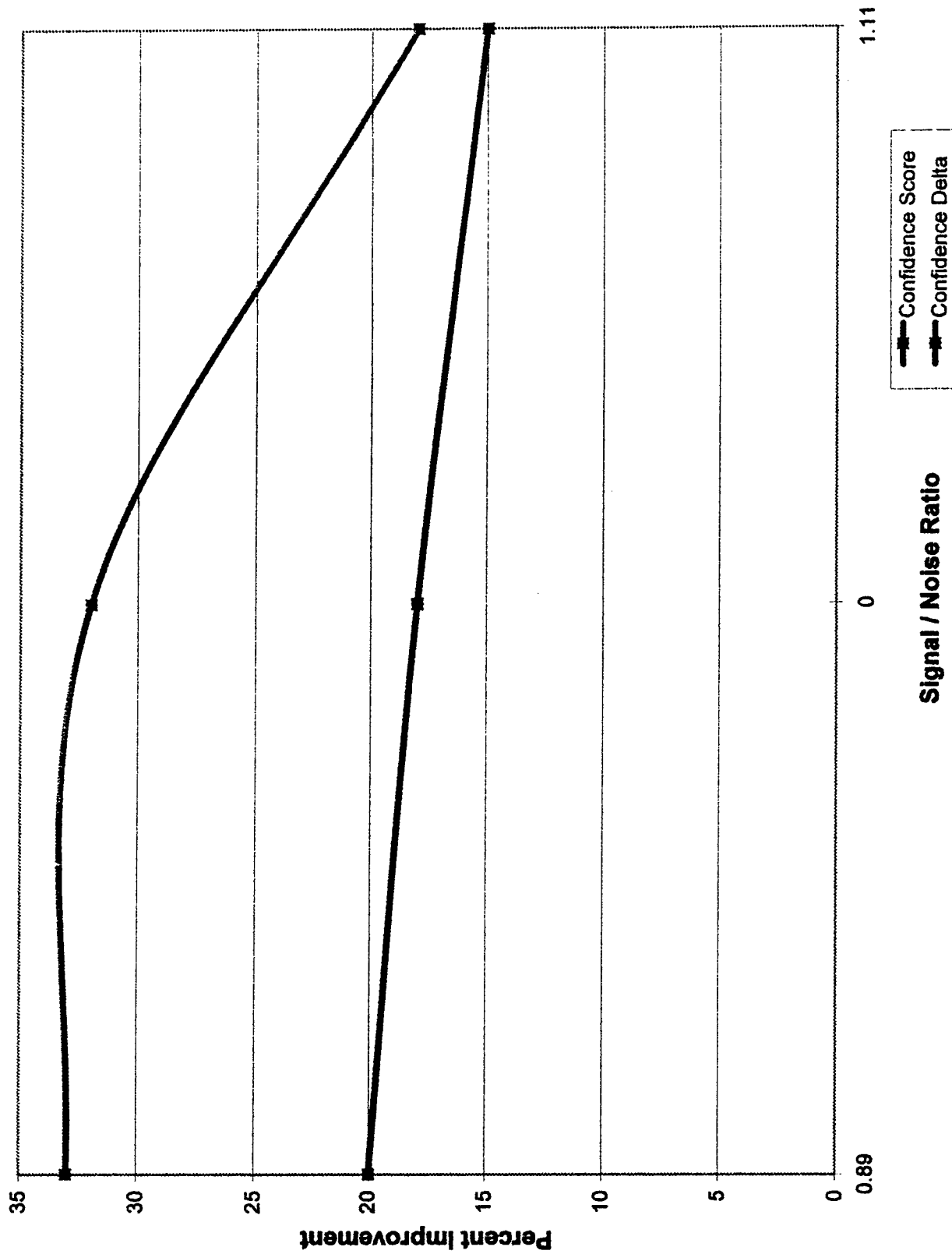
# **Confidence Delta vs. Noise Subtraction Coefficient For Various Speech Input Levels**



**FIGURE 7: CONFIDENCE DELTA WITH 90 dB NOISE @ 2 KHz**



**FIGURE 8: CONFIDENCE SCORE AND DELTA IMPROVEMENT  
FOR VARIOUS BACKGROUND NOISE FREQUENCIES**



**FIGURE 9: CONFIDENCE SCORE AND DELTA IMPROVEMENT  
FOR VARIOUS SIGNAL-TO-NOISE RATIOS**

# Confidence Scores vs. Speech Input Levels For Various Noise Subtraction Coefficients

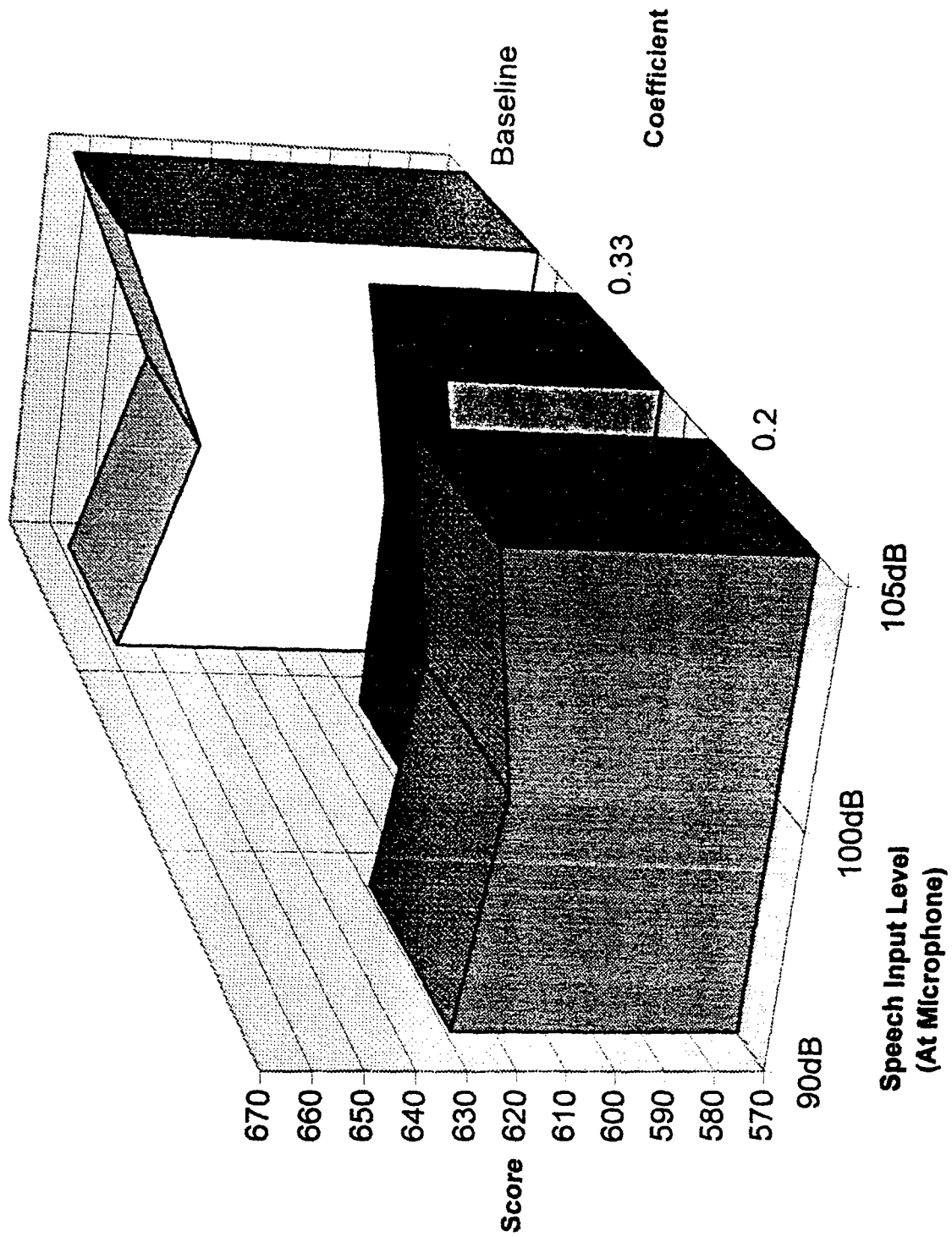


FIGURE 10: CONFIDENCE SCORES WITH 100 dB NOISE @ 500 Hz

# Confidence Delta vs. Speech Input Levels For Various Noise Subtraction Coefficients

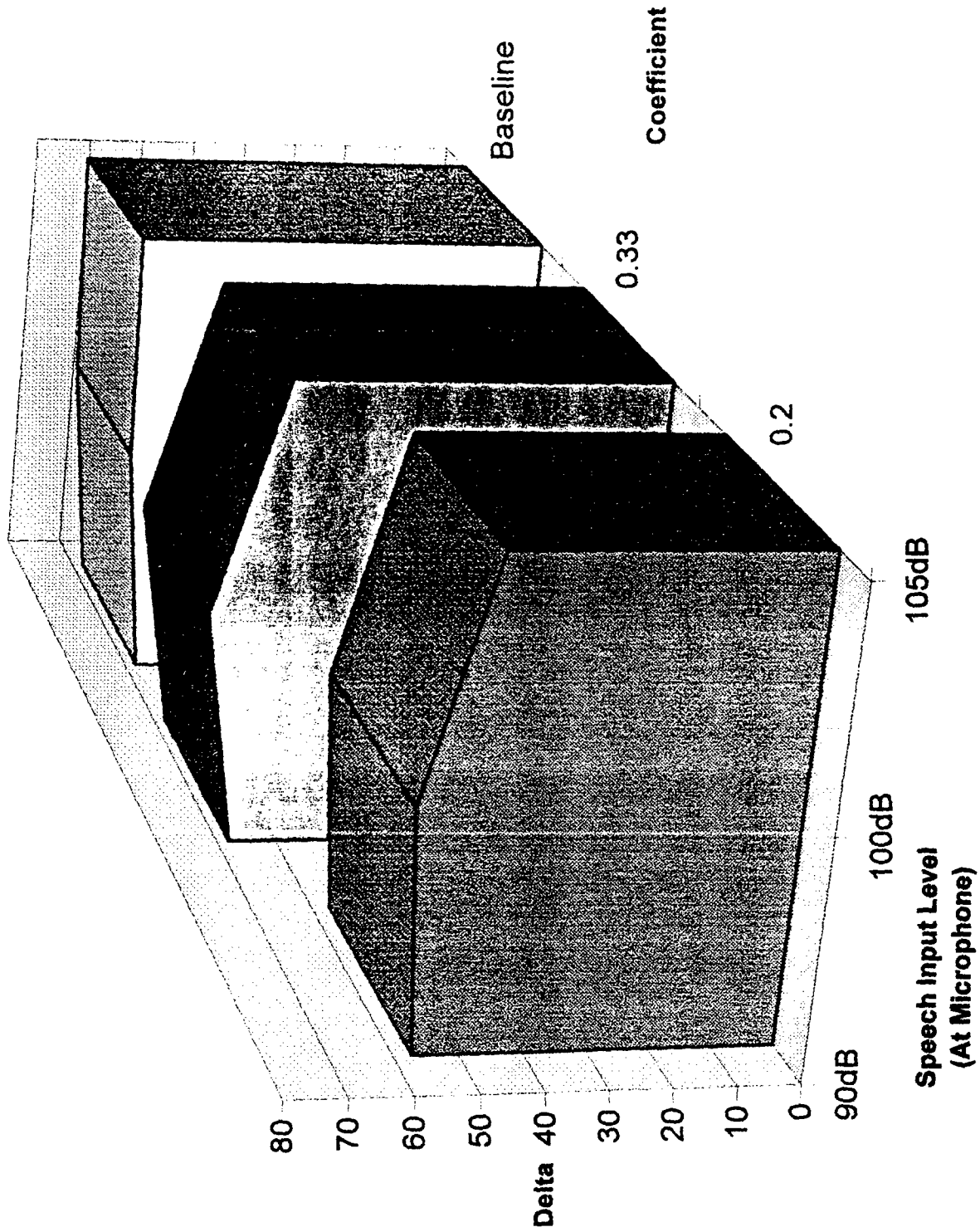


FIGURE 11: CONFIDENCE DELTA WITH 100 dB NOISE @ 500 Hz

# Confidence Scores vs. Speech Input Levels For Various Noise Subtraction Coefficients

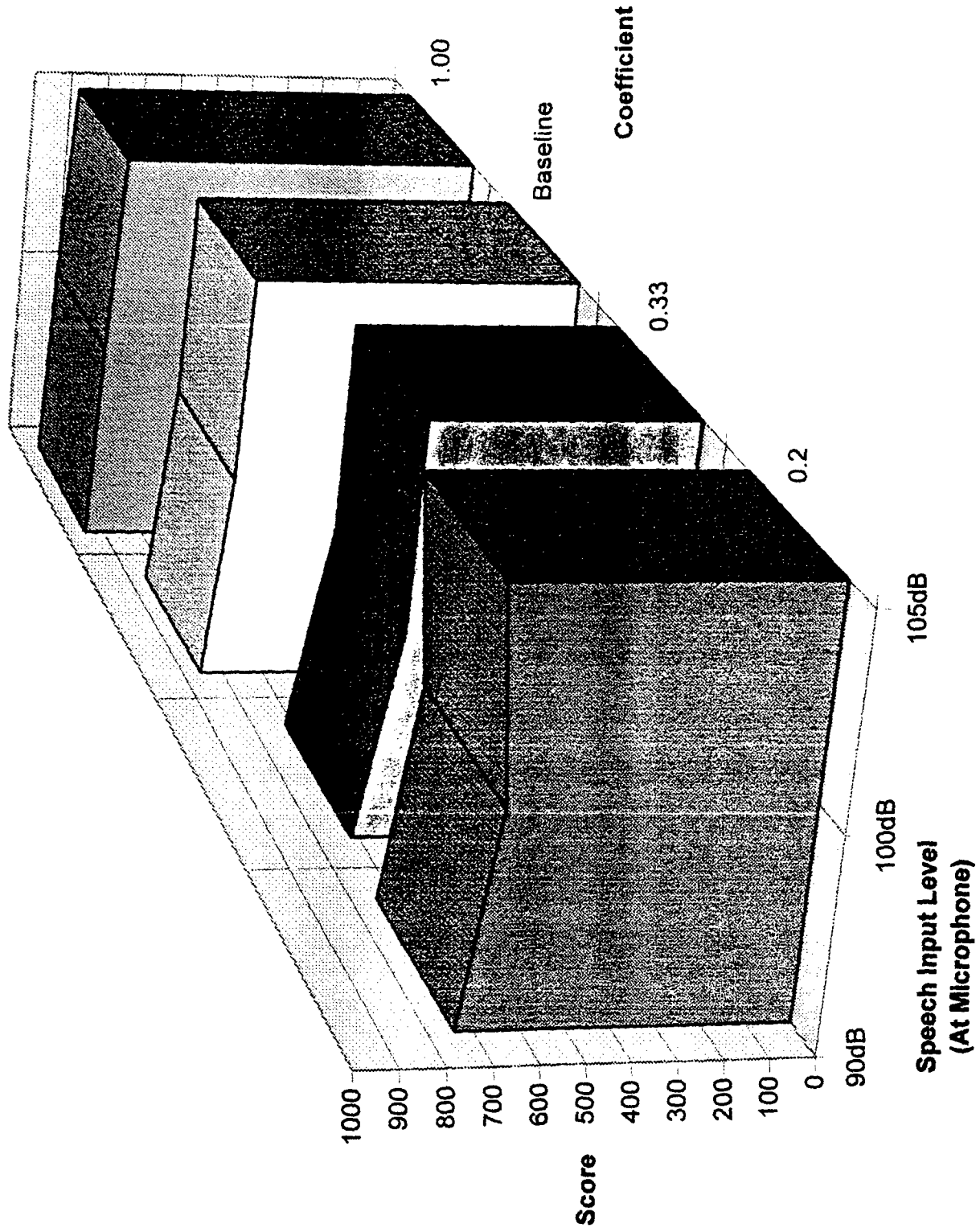


FIGURE 12: CONFIDENCE SCORES WITH 100 dB NOISE @ 1 KHz

# Confidence Delta vs. Speech Input Levels For Various Noise Subtraction Coefficients

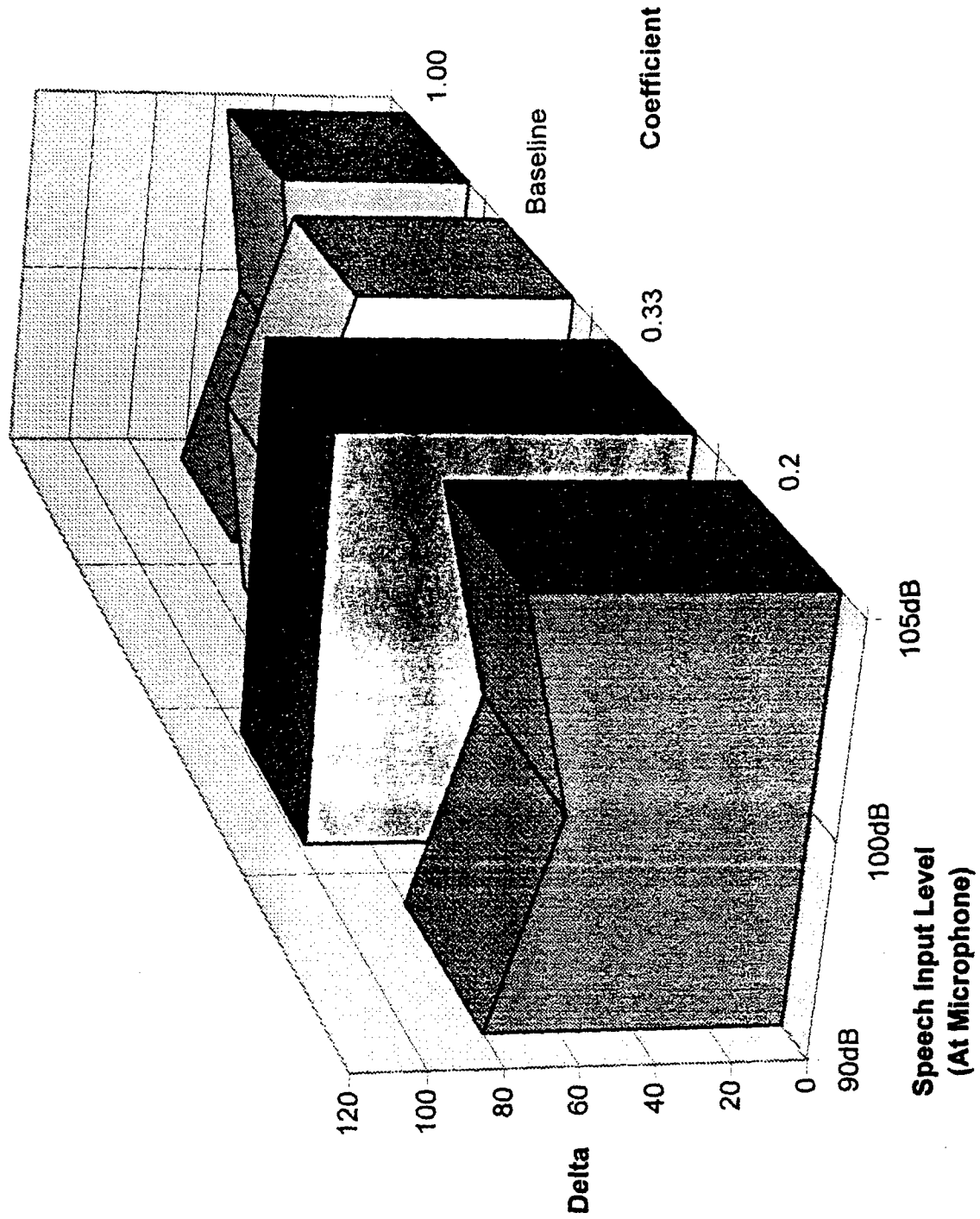


FIGURE 13: CONFIDENCE DELTA WITH 100 dB NOISE @ 1 KHz



# Confidence Scores vs. Speech Input Levels For Various Noise Subtraction Coefficients

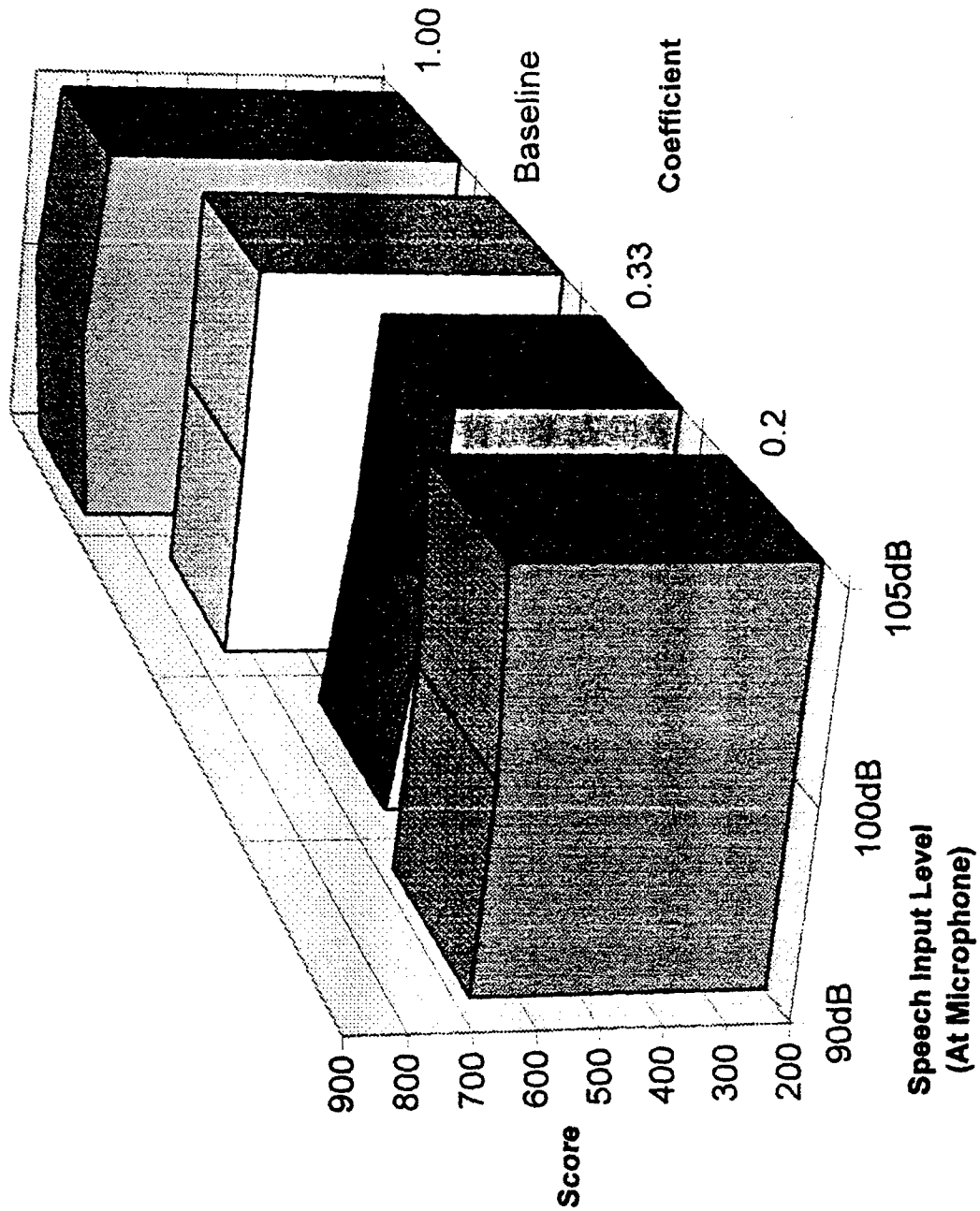
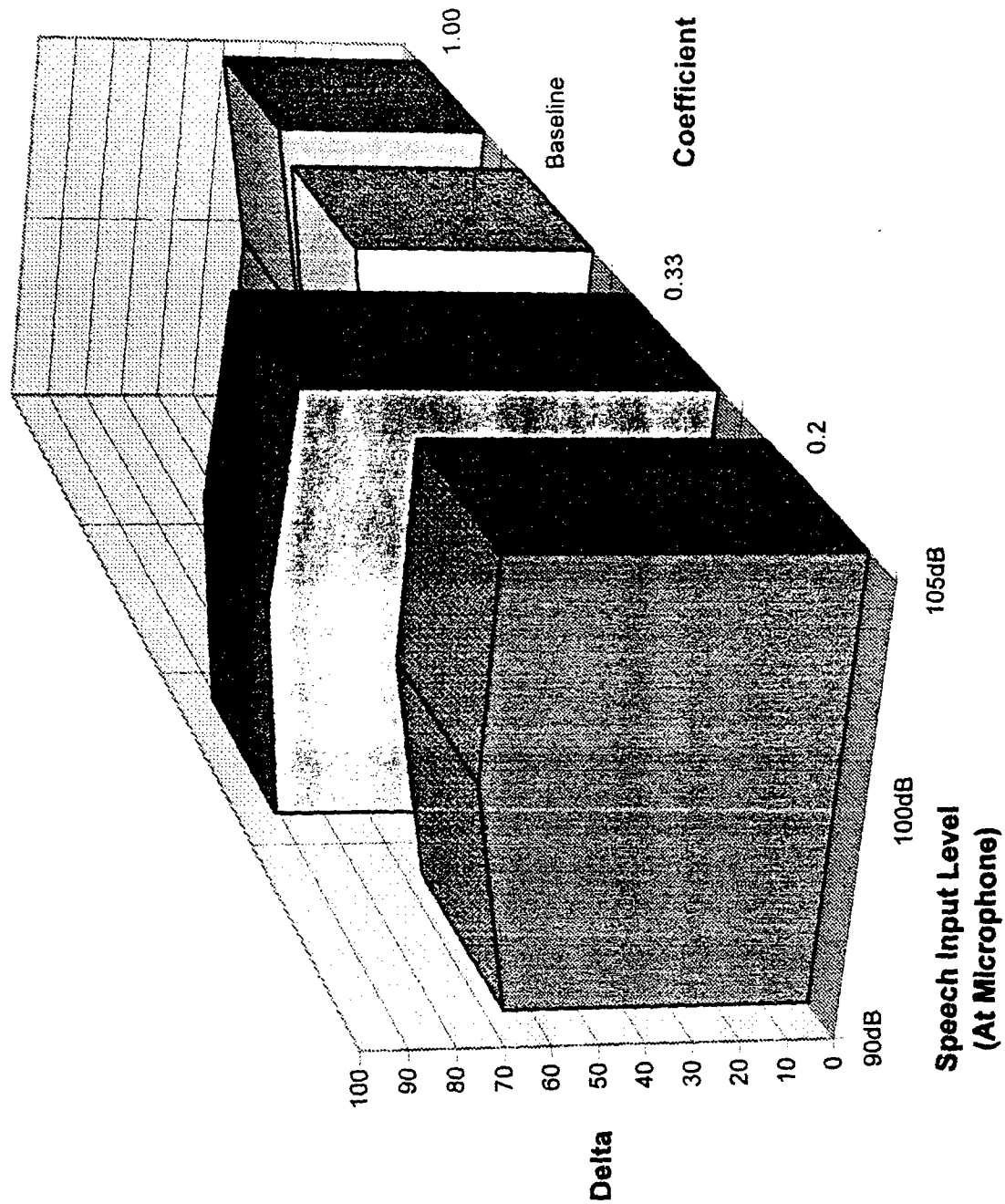


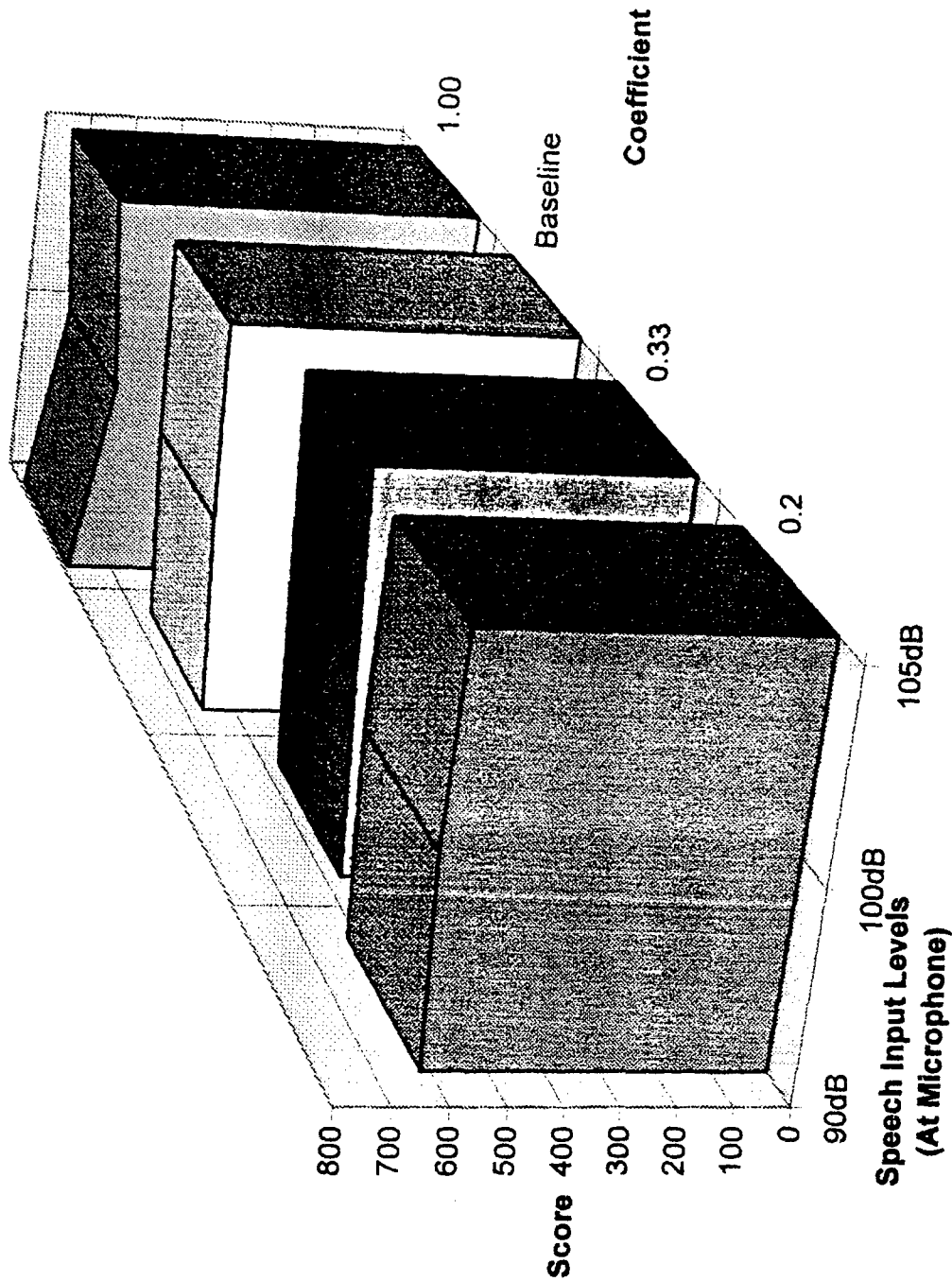
FIGURE 14: CONFIDENCE SCORES WITH 100 dB NOISE @ 2 KHz

# **Confidence Delta vs. Speech Input Levels For Various Noise Subtraction Coefficients**



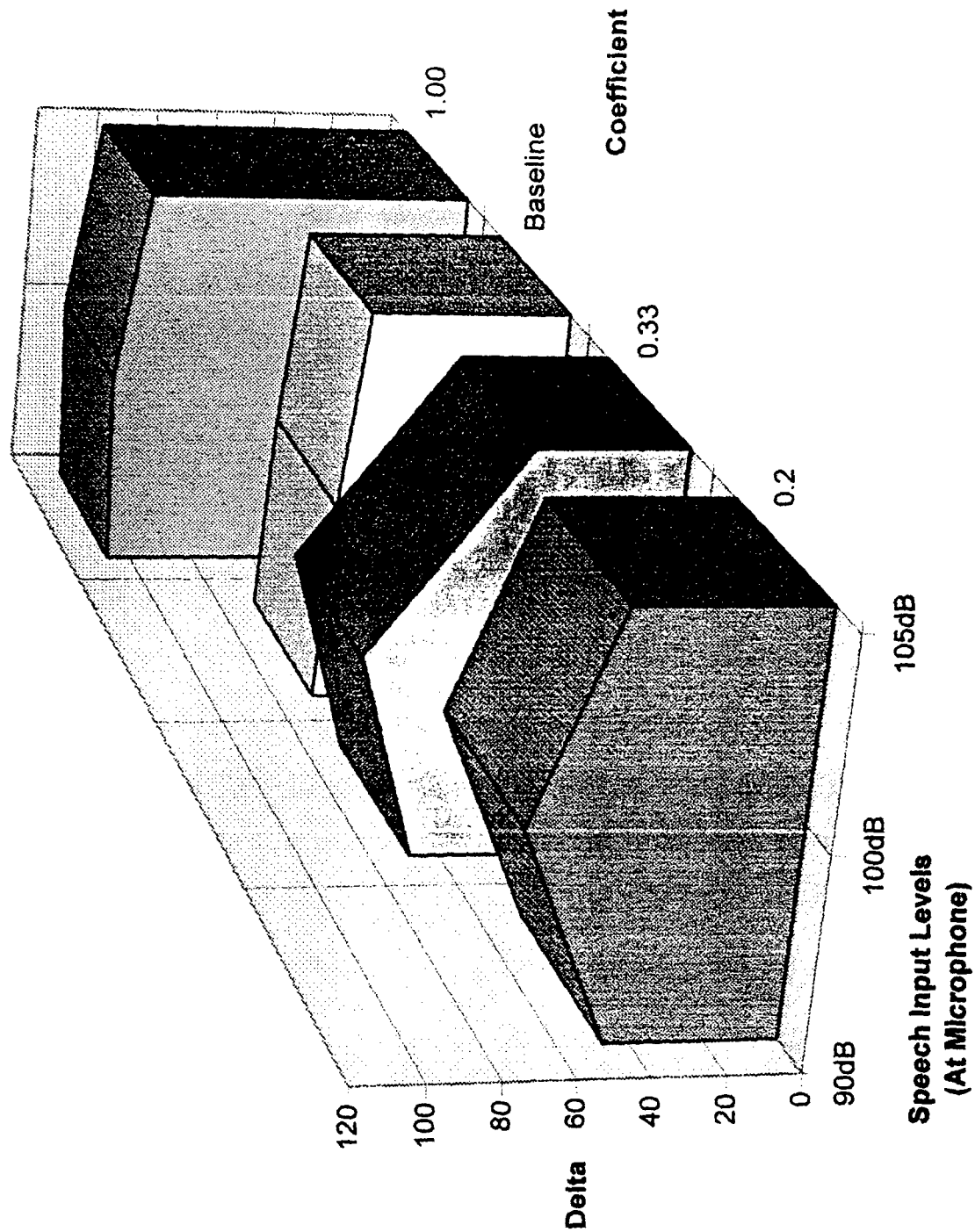
**FIGURE 15: CONFIDENCE DELTA WITH 100 dB NOISE @ 2 KHz**

# Confidence Scores vs. Speech Input Levels For Various Noise Subtraction Coefficients

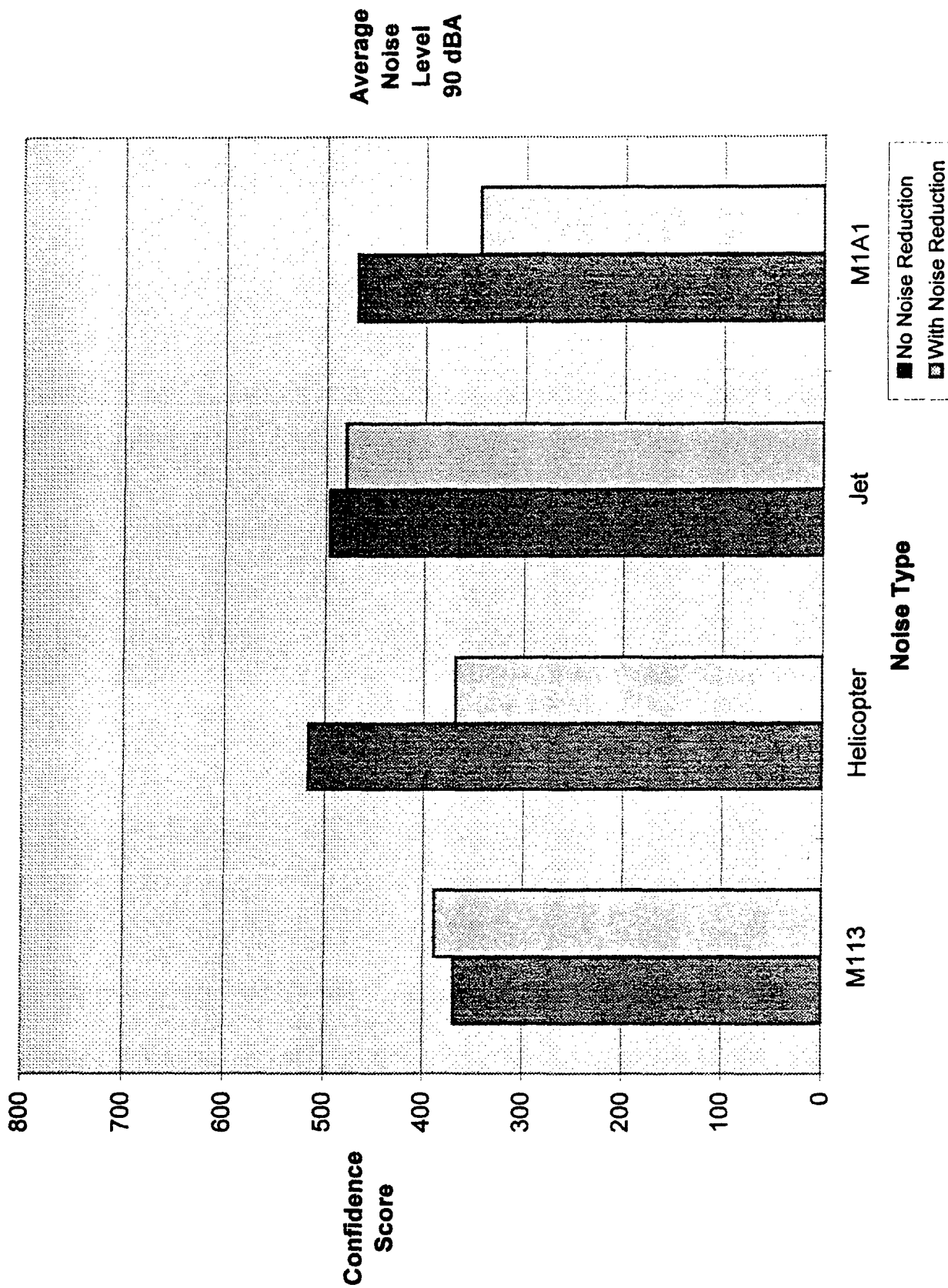


**FIGURE 16: CONFIDENCE SCORES WITH 100 dB  
M1A1 BACKGROUND NOISE**

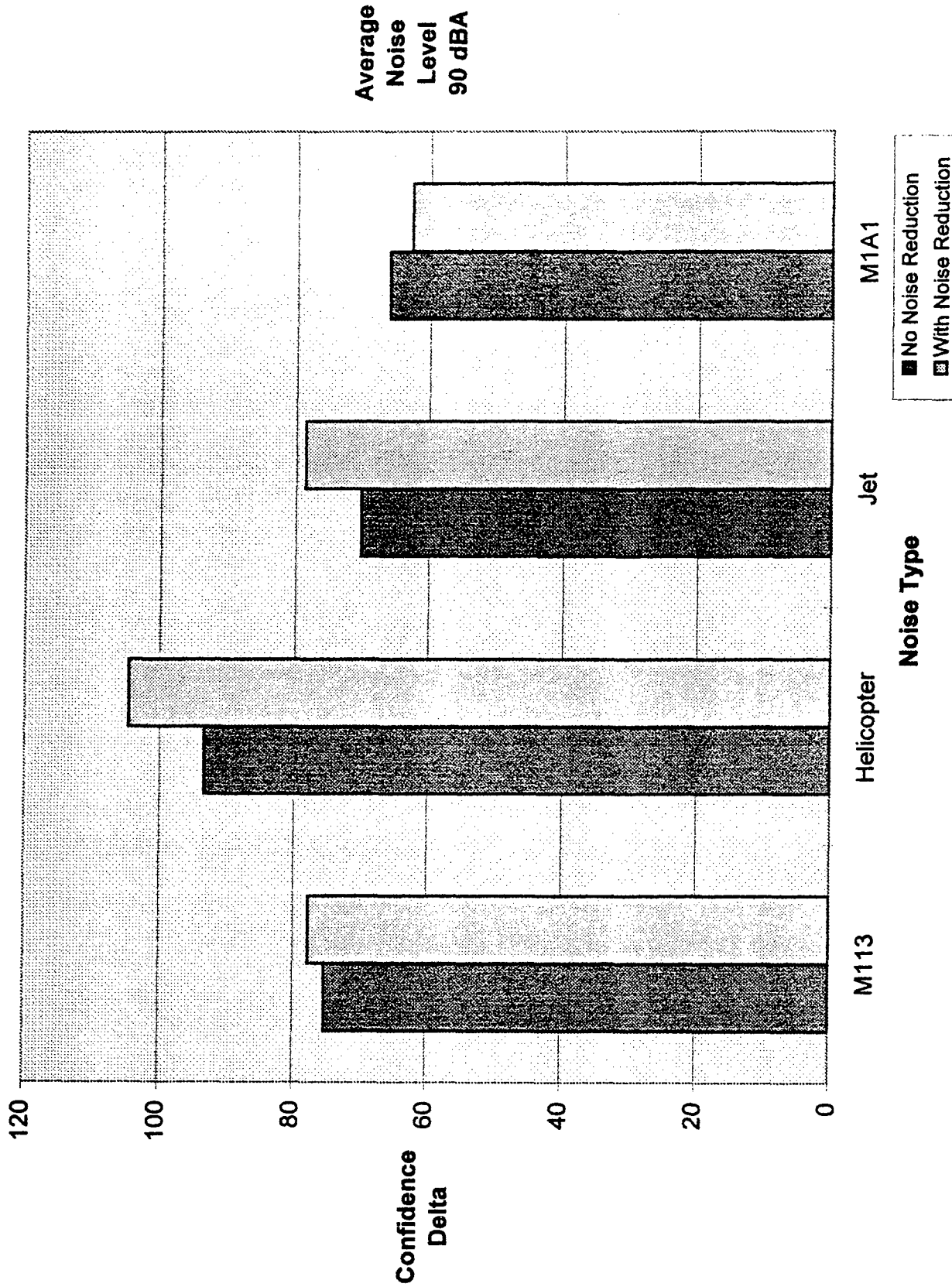
# Confidence Deltas vs. Speech Input Levels For Various Noise Subtraction Coefficients



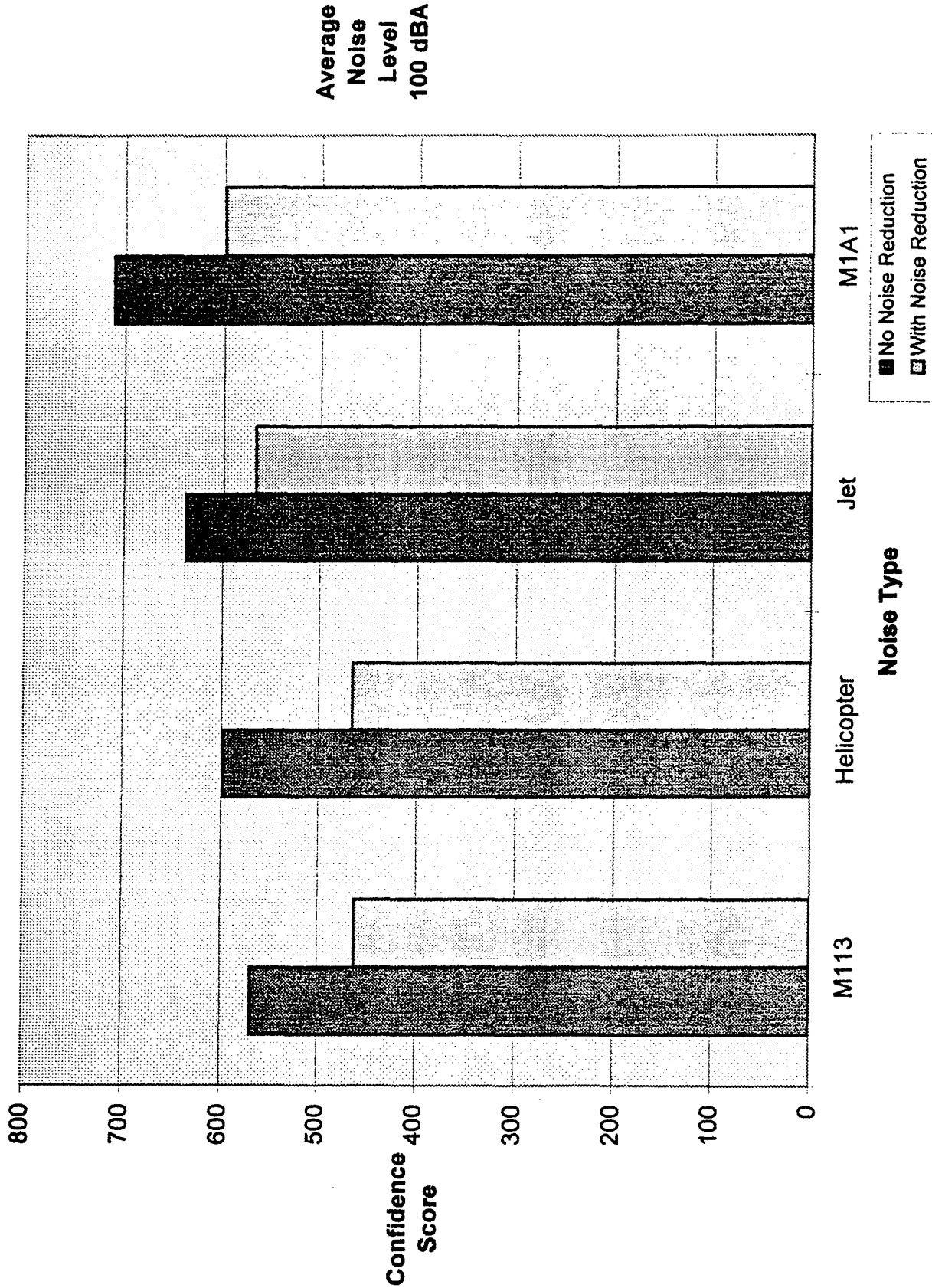
**FIGURE 17: CONFIDENCE DELTA WITH 100 dB  
M1A1 BACKGROUND NOISE**



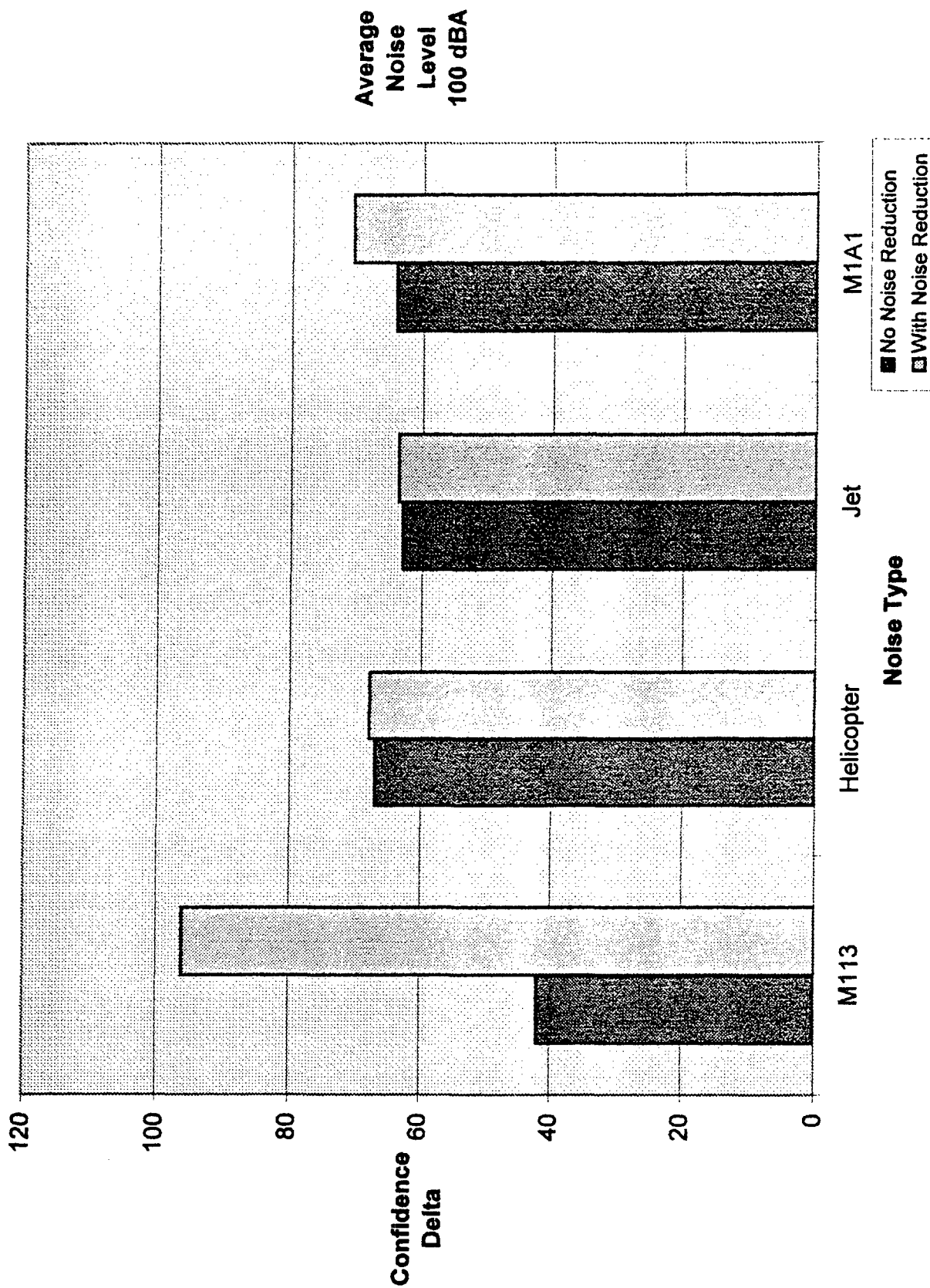
**FIGURE 18: CONFIDENCE SCORES FOR VARIOUS PRE-RECORDED  
FIELD NOISE TRIALS**



**FIGURE 19: CONFIDENCE DELTA FOR VARIOUS PRE-RECORDED  
FIELD NOISE TRIALS**



**FIGURE 20: CONFIDENCE SCORES FOR VARIOUS PRE-RECORDED  
FIELD NOISE TRIALS**



**FIGURE 21: CONFIDENCE DELTA FOR VARIOUS PRE-RECORDED  
FIELD NOISE TRIALS**



# Confidence Scores vs. Speech Input Levels For Various Noise Subtraction Coefficients

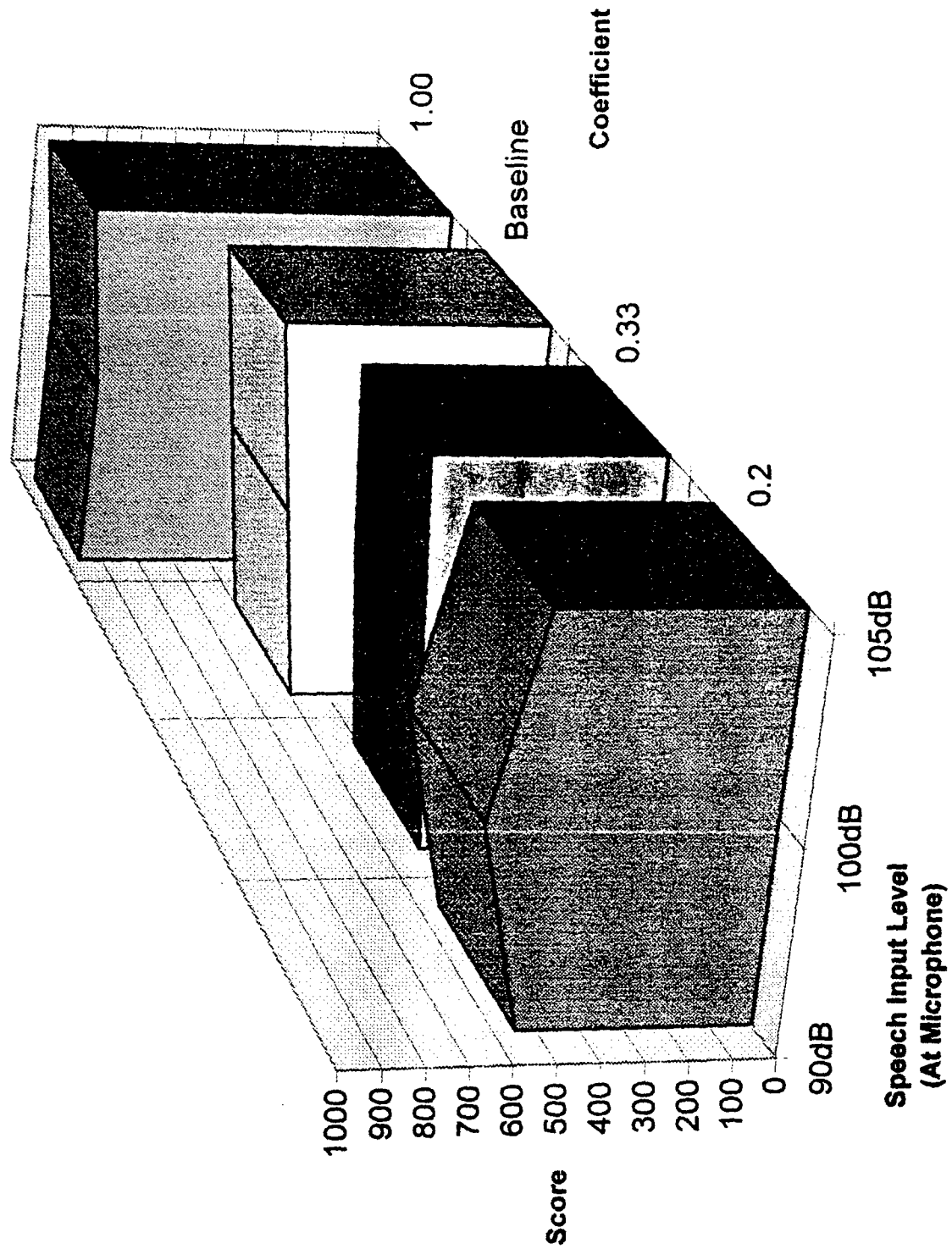


FIGURE 22: CONFIDENCE SCORES WITH 105 dB NOISE @ 500 Hz

# Confidence Delta vs. Speech Input Levels For Various Noise Subtraction Coefficients

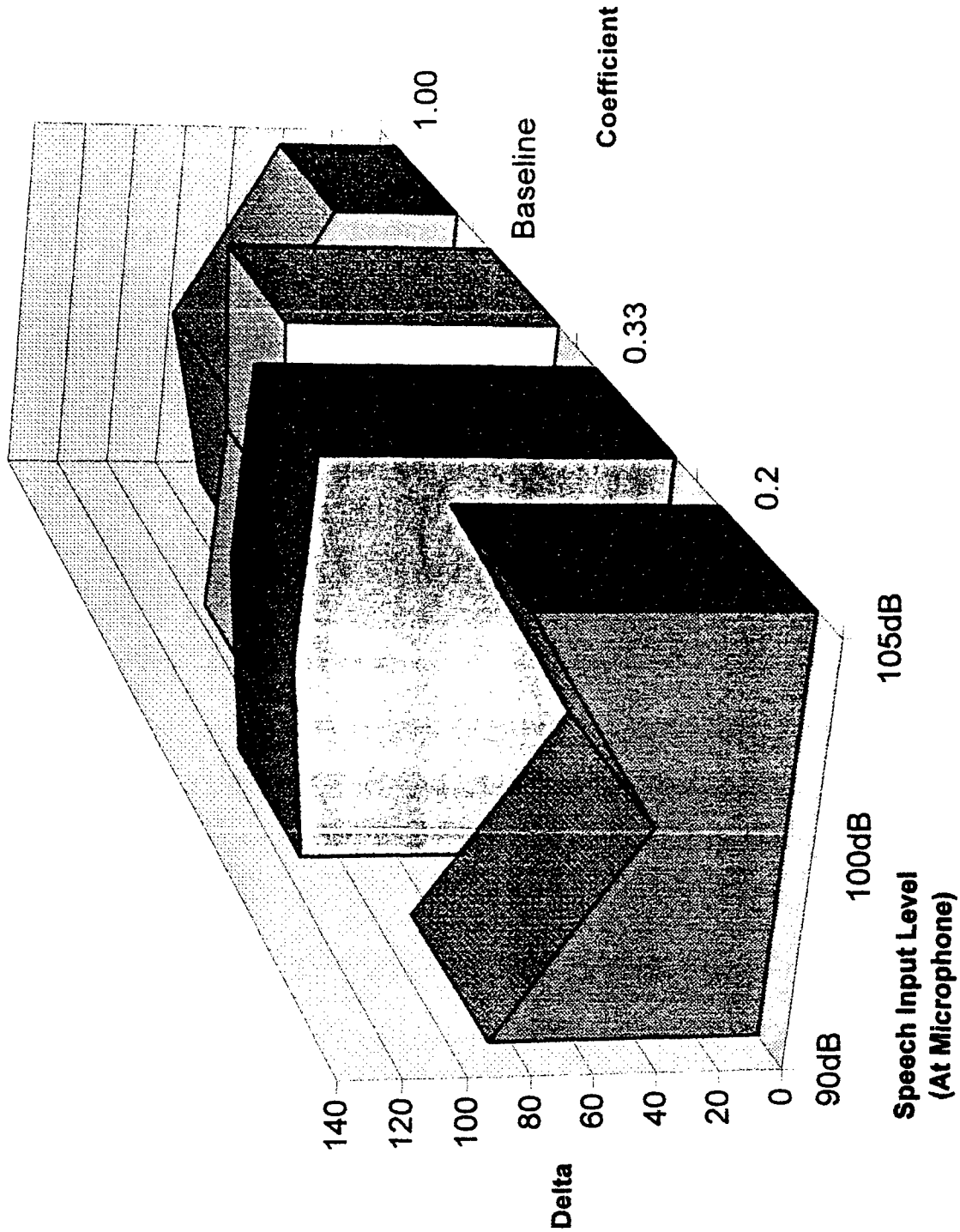


FIGURE 23: CONFIDENCE DELTA WITH 105 dB NOISE @ 500 Hz

# Confidence Scores vs. Speech Input Levels For Various Noise Subtraction Coefficients

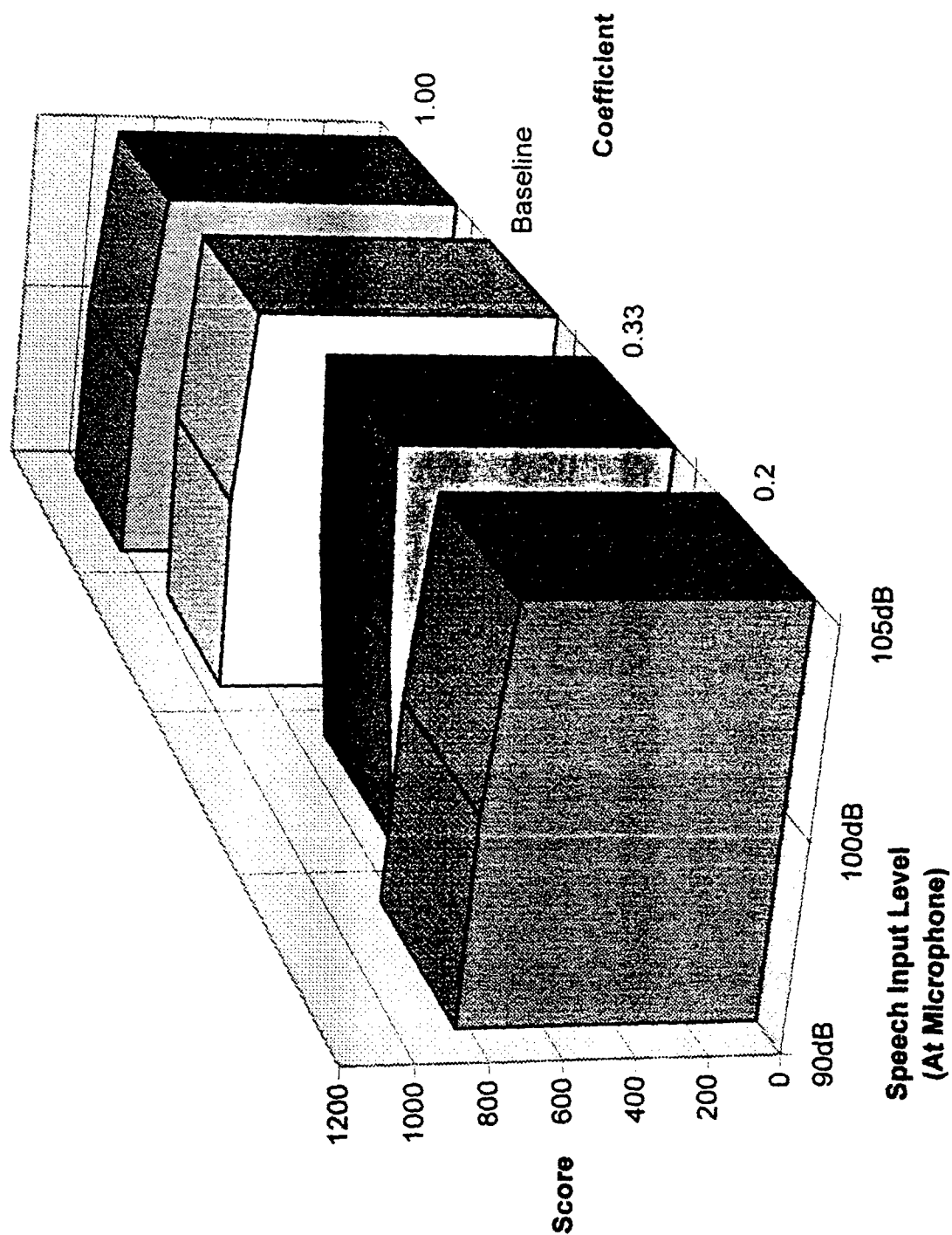


FIGURE 24: CONFIDENCE SCORES WITH 105 dB NOISE @ 1 KHz

# Confidence Delta vs. Speech Input Levels For Various Noise Subtraction Coefficients

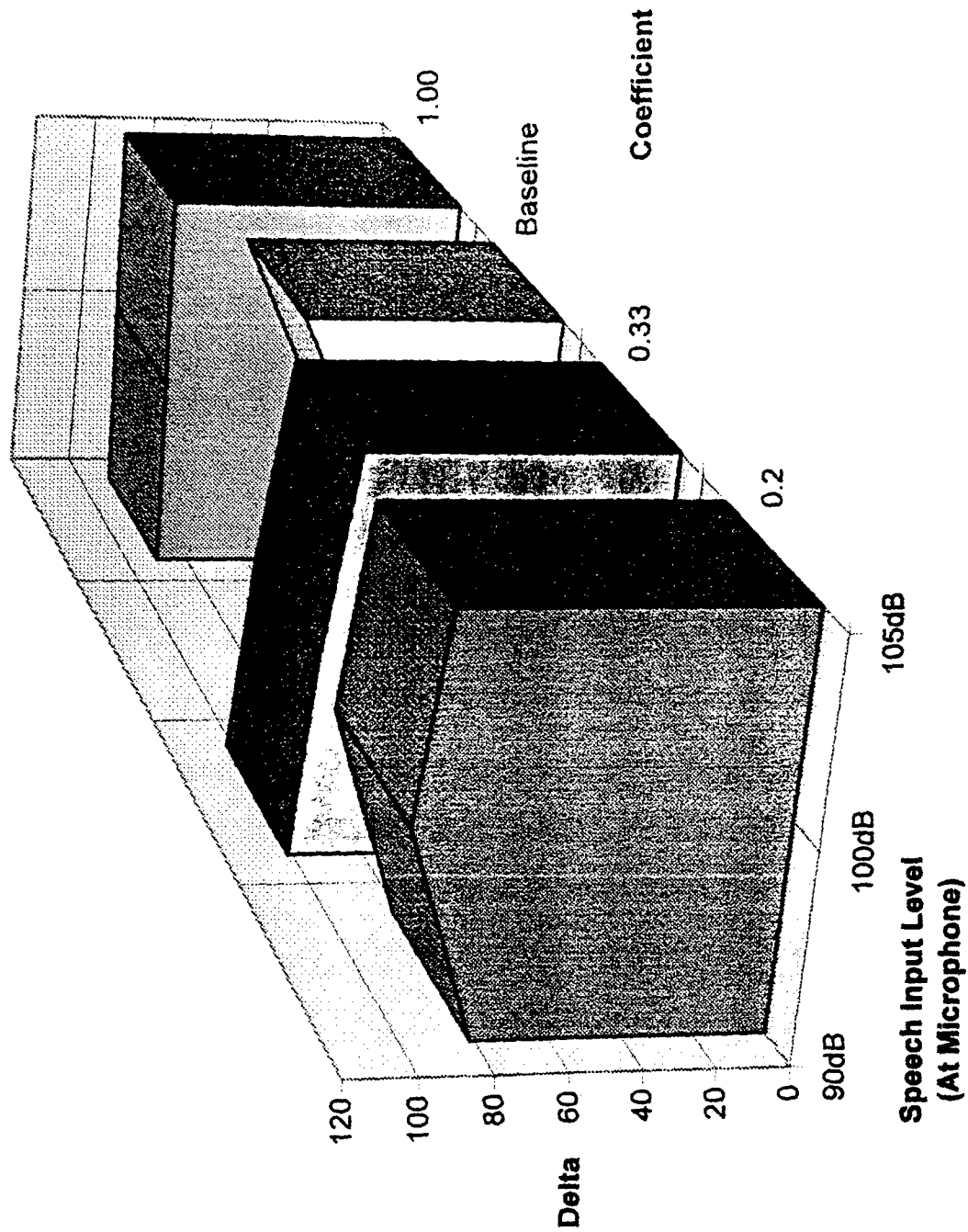
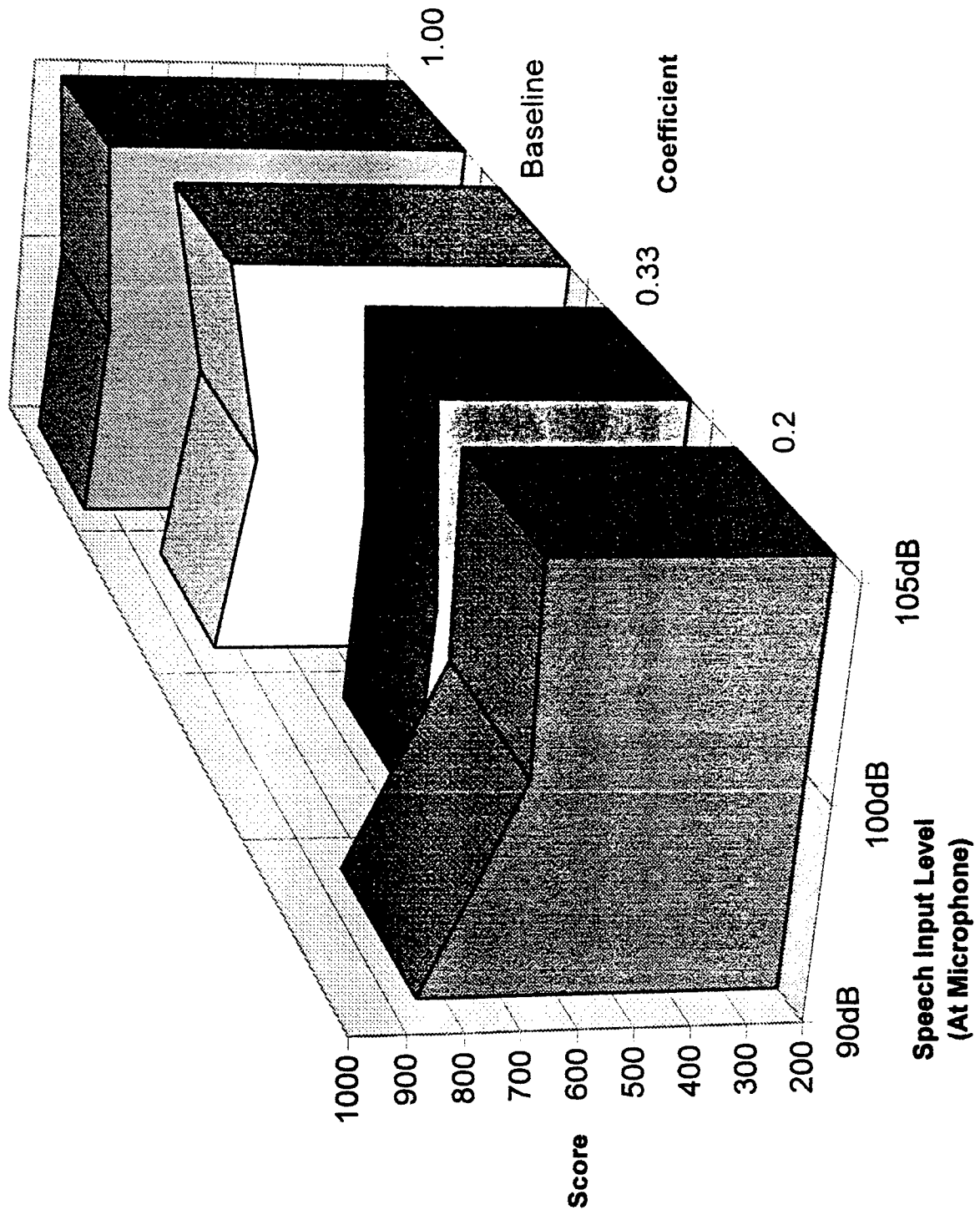


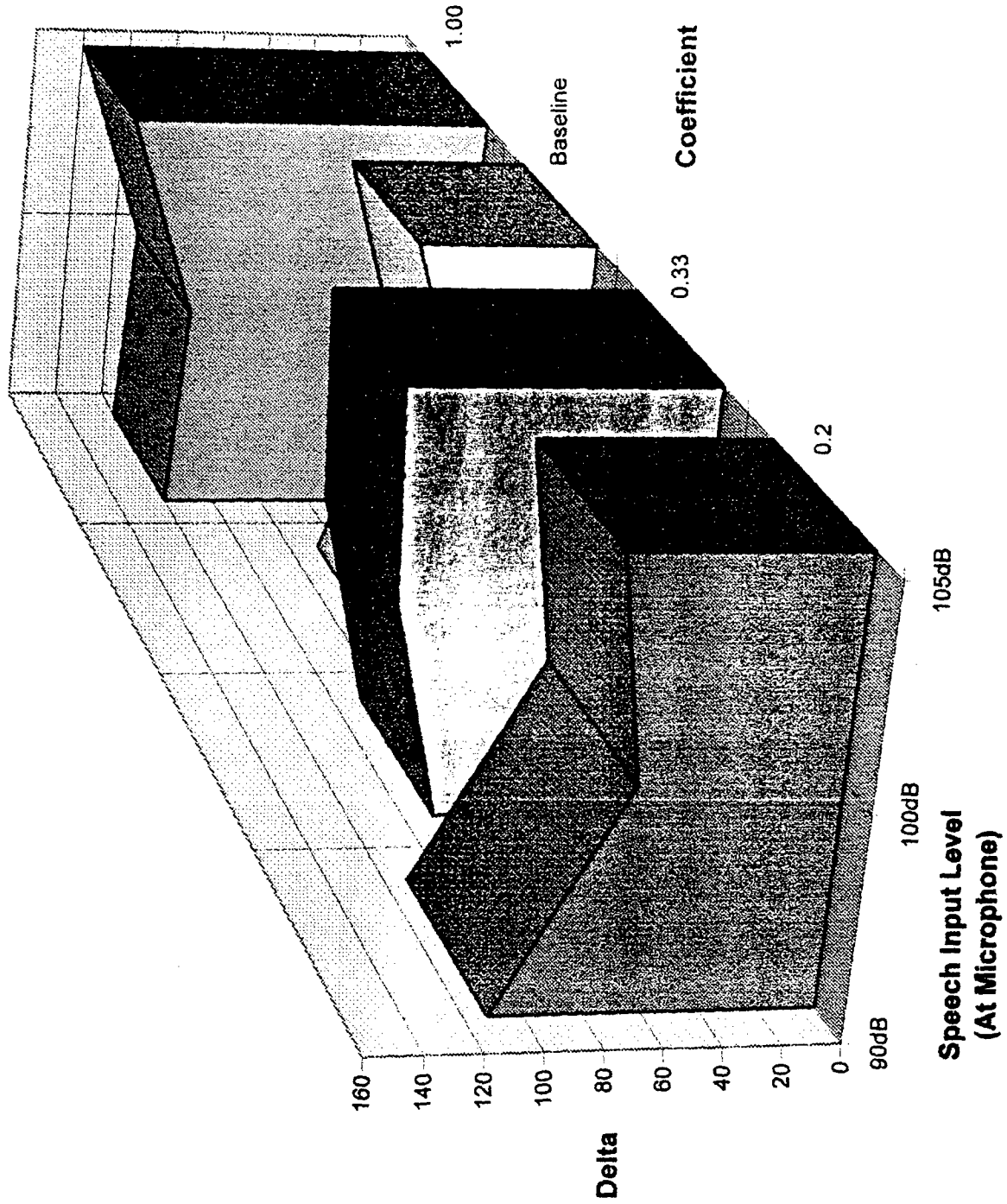
FIGURE 25: CONFIDENCE DELTA WITH 105 dB NOISE @ 1 KHz

### Confidence Scores vs. Speech Input Levels For Various Noise Subtraction Coefficients



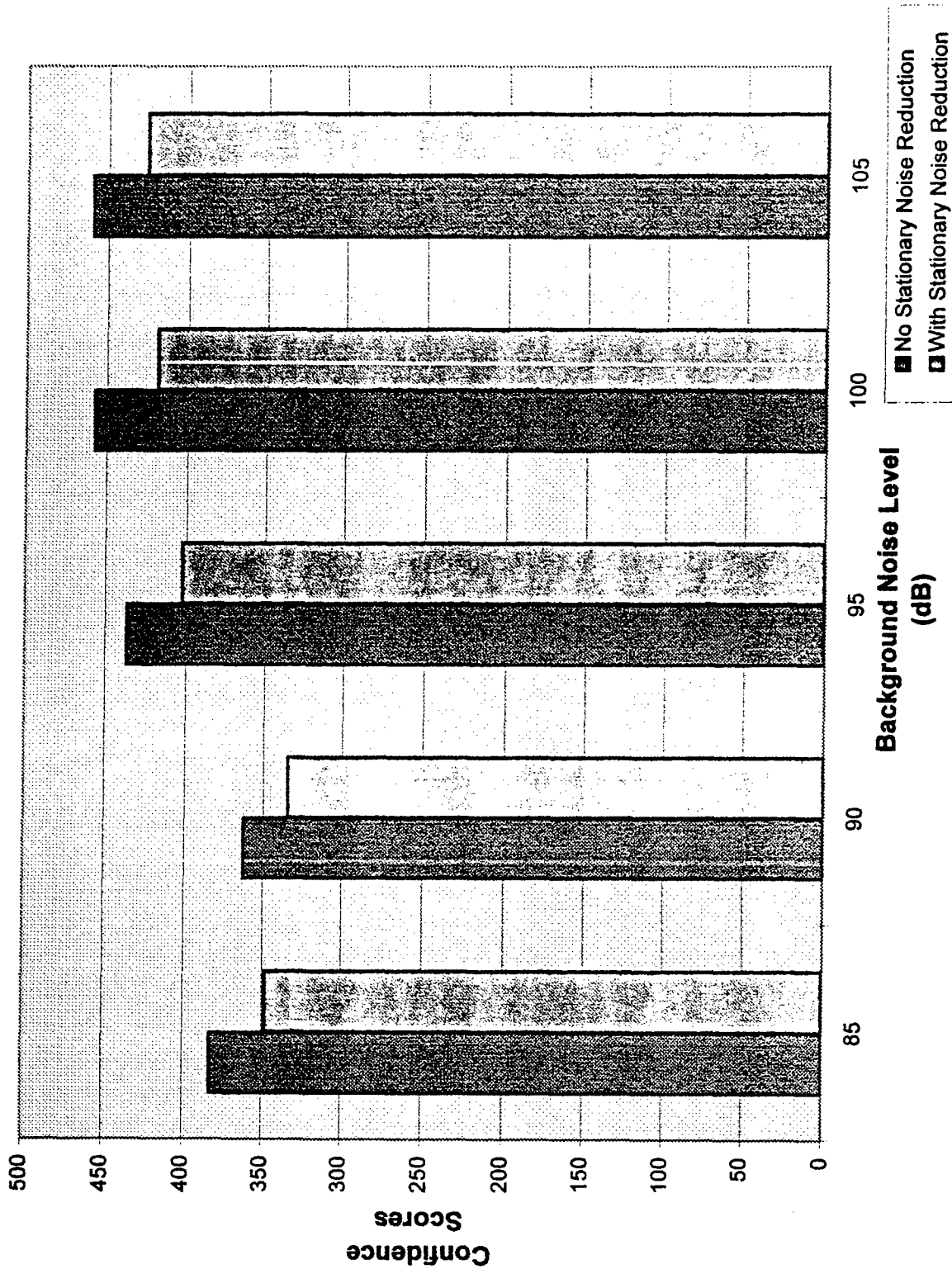
**FIGURE 26: CONFIDENCE SCORES WITH 105 dB NOISE @ 2 KHz**

# **Confidence Delta vs. Speech Input Levels For Various Noise Subtraction Coefficients**



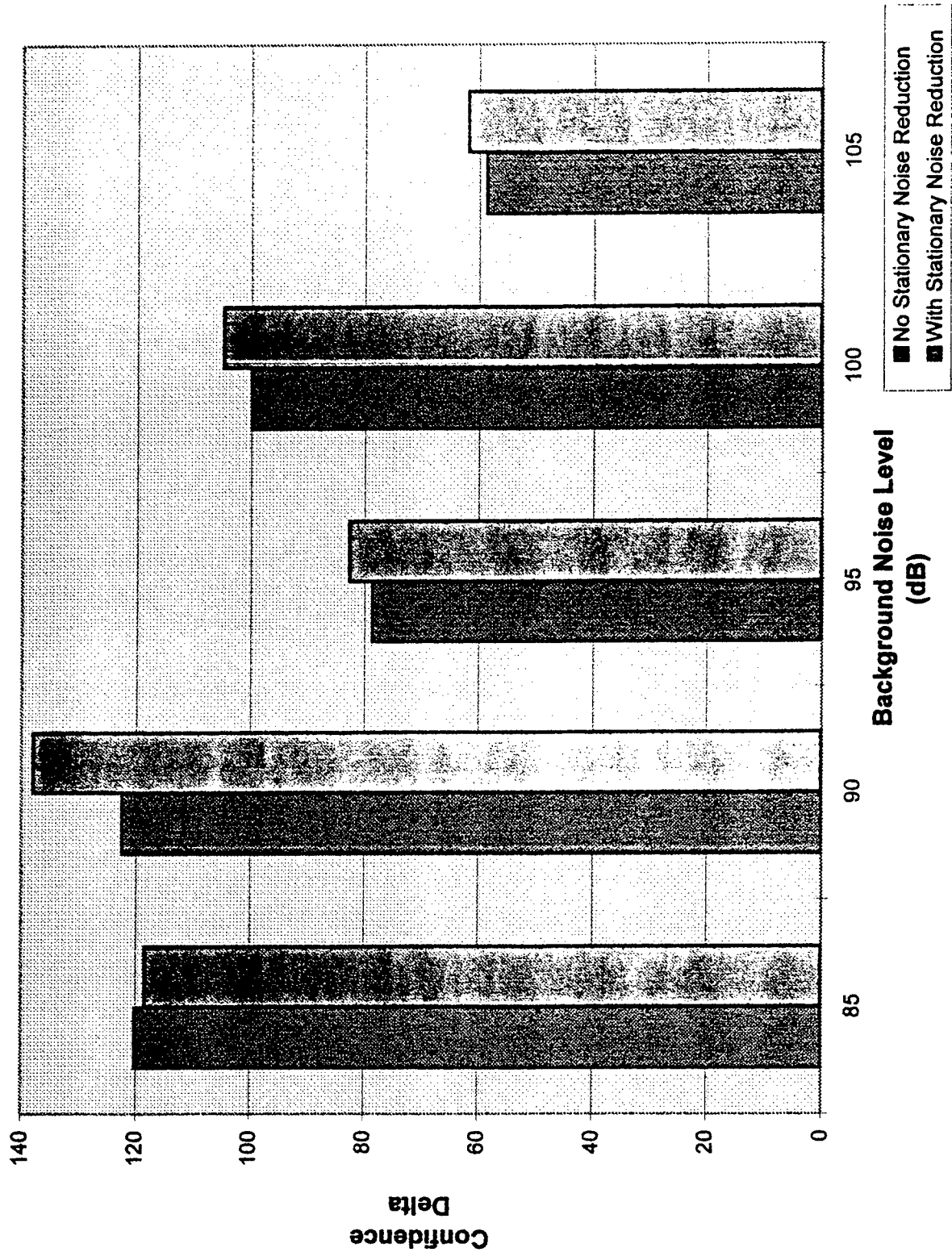
**FIGURE 27: CONFIDENCE DELTA WITH 105 dB NOISE @ 2 KHz**

# Confidence Scores vs. Tactical Background Noise Level



**FIGURE 28: SCAD CONFIDENCE SCORE COMPARISON WITH STATIONARY NOISE FEATURE EXTRACTION**

# Confidence Delta vs. Tactical Background Noise Levels



**FIGURE 29: SCAD CONFIDENCE DELTA COMPARISON WITH STATIONARY NOISE FEATURE EXTRACTION**





DEPARTMENT OF THE ARMY  
US ARMY MEDICAL RESEARCH AND MATERIEL COMMAND  
504 SCOTT STREET  
FORT DETRICK, MARYLAND 21702-5012

REPLY TO  
ATTENTION OF:

MCMR-RMI-S (70-1y)

4 Dec 02

MEMORANDUM FOR Administrator, Defense Technical Information  
Center (DTIC-OCA), 8725 John J. Kingman Road, Fort Belvoir,  
VA 22060-6218


SUBJECT: Request Change in Distribution Statement

1. The U.S. Army Medical Research and Materiel Command has reexamined the need for the limitation assigned to technical reports written for this Command. Request the limited distribution statement for the enclosed accession numbers be changed to "Approved for public release; distribution unlimited." These reports should be released to the National Technical Information Service.

2. Point of contact for this request is Ms. Kristin Morrow at DSN 343-7327 or by e-mail at Kristin.Morrow@det.amedd.army.mil.

FOR THE COMMANDER:

Encl

  
PHYLLIS M. RINEHART  
Deputy Chief of Staff for  
Information Management

ADB218773	ADB229914
ADB223531	ADB229497
ADB230017	ADB230947
ADB223528	ADB282209
ADB231930	ADB270846
ADB226038	ADB282266
ADB224296	ADB262442
ADB228898	ADB256670
ADB216077	
ADB218568	
ADB216713	
ADB216627	
ADB215717	
ADB218709	
ADB216942	
ADB216071	
ADB215736	
ADB216715	
ADB215485	
ADB215487	
ADB220304	
ADB215719	
ADB216072	
ADB222892	
ADB215914	
ADB222994	
ADB216066	
ADB217309	
ADB216726	
ADB216947	
ADB227451	
ADB229334	
ADB228982	
ADB227216	
ADB224877	
ADB224876	
ADB227768	
ADB228161	
ADB229442	
ADB230946	
ADB230047	
ADB225895	
ADB229467	
ADB224342	
ADB230950	
ADB227185	
ADB231856	